CLONING, EXPRESSION, AND PURIFICATION OF THE *DROSOPHILA MELANOGASTER* DOSAGE COMPENSATION COMPLEX CHROMODOMAINS AND THEIR *HOMO SAPIENS* ORTHOLOGUES.

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Masters of Science

In the Department of Biochemistry

University of Saskatchewan

Saskatoon

By

Andrew James Welham

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a postgrduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or part, for scholary purposes may be granted by the professors who supervised the work, or in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and the University of Saskatchewan in any cholary use which may be made of any materials in my thesis.

Request for permission to copy or make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Biochemistry
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5E5

Abstract

Sexual differentiation is a fundamental characteristic of all eukaryotes, dictating sex-specific morphology, physiology and behavior. Diploid organisms with heteromorphic sex chromosomes (XX or XY) require regulatory compensation of the X chromosome to maintain correct levels of genetic expression between the sexes, a process termed sex-specific dosage compensation (SSDC). The fruit fly, *Drosophila melanogaster* dosage compensates by upregulating transcription of most X-linked genes two-fold. Associated with this two-fold up regulation is the male-specific lethal (MSL) complex, a RNA-protein complex comprised of at least five known proteins; MSL1, MSL2, MSL3, males absent on the first (MOF), and maleless (MLE) and two non-translated RNA molecules; roX1 (RNA on the X chromosome) and roX2. The complex modulates the chromatin structure of the male X chromosome via acetylation of H4K16. MOF and MSL3 both exhibit an N-terminal chromodomain, whose function is unclear. The MSL3 chromodomain has been suggested to bind H3K36Me_3. Chromodomains are a paradigm of how a single structural fold has evolved in diverse proteins to bind distinct targets. Chromodomains are common to nuclear regulators, and bind diverse targets including histones, DNA, and RNA. They function as recognition motifs of histone post-translational modifications and facilitate the translation of the histone code into a distinct local chromatin structure via recruiting the appropriate chromatin modulating machinery.

The goal of this research is to determine the structure of the *D. melanogaster* MOF and MSL3 chromodomains by X-ray crystallographic and/or nuclear magnetic resonance techniques, to advance our understanding of the structural characteristics of these diverse domains. Here we report the cloning and reproducible expression and purification of the *D. melanogaster* MOF and MSL3 chromodomains and their *Homo sapiens* orthologues. The *D. melanogaster* MOF chromodomain, whose NMR structure was published during this research, has been crystallized. Attempts to solve the crystal structure by molecular replacement, multiple-wavelength anomalous dispersion, and single-wavelength isomorphous replacement are reported.

Table of Contents

List of Tables

List of Figures

List of Equations

List of Abbreviations

ACF	ATP-utilizing chromatin assembly and remodeling factor
Amp	Ampicillin
APS	Ammonium persulphate
ATP	Adenosine triphosphate
BRM	Brahma
BSA	Bovine serum albumin
CCD	Charge-coupled device
CHD1	Chromo-ATPase/helicase-DNA binding domain 1
CHES	2-(Cyclohexylamino)ethanesulfonic acid
CHRAC	Chromatin accessibility complex
CLS	Canadian Light Source
Da	Dalton
DC	Dosage compensation
DCC	Dosage compensation complex
ddH_2O	Double-distilled H_2O
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EAF3	Esa1 associated factor 3
EDTA	Ethylenediaminetetraacetic acid
ESA1	Essential Sas-related acetyltransferase 1
(F_H)	Heavy atom structure factor
(F_P)	Native structure factor
(F_{PH})	Structure factor of the heavy atom derivative
FRF	Fast rotation function
FTF	Fast translation function
GST	Glutathione S-transferase
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid

HP1	Heterochromatin protein 1
HSQC	Heteronuclear single quantum coherence
IPTG	Isopropyl β-D-thiogalactopyranoside
ISWI	Imitation switch
LB	Luria Bertani Broth
LLG	Log-likelihood gain
MAD	Multi-wavelength anomalous dispersion
mAU	Milli-absorbance units
MIR	Multiple isomorphous replacement
MLE	Male-less
MOF	Males absent on the first
MRG	MORF4-related gene
mS	Milli-sieverts
MSL	Male-specific lethal
MW	Molecular weight
NCP	Nucleosome core particle
NDC	Network dosage compensation
NMR	Nuclear magnetic resonance
NURF	Nucleosome remodeling factor
O/N	Over night
OD	Optical Density
ORF	Open reading frame
PBS	Phosphate buffered saline
PC	Polycomb
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
Pfu	*Pyrococcus furiosus*
PMSF	Phenylmethylsulphonyl fluoride
PTM	Post-translational modification
RD	Replication dependent
RI	Replication independent

rms	Root-mean-squared
rmsd	Root-mean-square difference
RNA	Ribonucleic acid
roX	RNA on the X
SAD	Single anomalous dispersion
SAGA	Spt-Ada-Gcn5-Acetyltransferase
SCF	Supercoiling factor
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate Polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SIR	Single isomorphous replacement
SIRAS	Single isomorphous replacement with anomalous scattering
SLIK	SAGA-like
SMN	Survival of Motor Neuron
SSDC	Sex-specific dosage compensation
SSSC	Saskatchewan Structural Sciences Center
SWI/SNF	Switch of mating type/Sucrose-nonfermenting
SXL	Sex-lethal
TAE	Tris-Acetate-EDTA
TAFI	Thrombin-activatable fibrinolysis inhibitor
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TE	Tris-EDTA
TEMED	Tetramethylethylenediamine
Tris-HCl	Tris(hydroxymethyl)aminomethane hydrochloride
TRL	Trithorax-like

1.0 Literature Review

1.1 Introduction

Epigenetic regulation in biology refers to the regulation of gene expression by a mechanism that does not directly alter the genetic code. Epigenetic regulation involves DNA methylation, chromatin remodeling and post-translational modification of histone tails. Dosage compensation, the equalization of X-linked gene expression between males and females is a paradigm of epigenetic regulation. The fruit fly, *Drosophila melanogaster* dosage compensates by up-regulating transcription of most male X-linked genes two-fold. Associated with this two-fold up regulation is the male-specific lethal (MSL) complex, a RNA-protein complex comprised of at least five known proteins; MSL1, MSL2, MSL3, males absent on the first (MOF), and maleless (MLE) and two non-translated RNA molecules; roX1 (RNA on the X chromosome) and roX2. The MSL protein-RNA complex modulates the chromatin structure of the male X chromosome, presumably via acetylation of K16 of histone H4. MOF and MSL3 both exhibit an N-terminal chromodomain. Numerous chromatin modification complexes exhibit chromodomains, where they function as a recognition module of histone post-translational modifications, facilitating chromatin interactions.

This research aims to structurally characterize the MSL3 and MOF chromodomains of the *D. melanogaster* dosage compensation complex, and their human orthologues. Structural investigation will include X-ray crystallographic and NMR techniques, which will involve crystallizing and isotopically labeling of the recombinant proteins, respectively. Solution of the structures will advance our understanding of the structural characteristics of these diverse domains. The following literature review places this research in context, and will focus specifically on epigenetic regulation and how it pertains to dosage compensation in *D. melanogaster*. The literature review will begin with a discussion of sexual differentiation and determination, and how this dictates dosage compensation. This is followed by the introduction of chromatin structure and how epigenetic regulation, and ultimately dosage compensation arises via chromatin post-translational modification, remodeling, and histone variation. This sets the stage for the introduction of the *D. melanogaster* dosage compensation complex, its components and mechanism. The structure and function of chromodomains is then summarized,

with specific focus on the MOF and MSL3 chromodomains of the *D. melanogaster* dosage compensation complex.

1.2 Sex

Sexual differentiation is a fundamental characteristic of higher eukaryotes, dictating sex-specific morphology, physiology and behavior. Interestingly, the mechanism of sexual differentiation between organisms is not widely conserved. A number of different genetic and epigenetic systems have evolved to generate sexual dimorphism. Diploid organisms with heteromorphic sex chromosomes (XX or XY) require regulatory compensation of the X chromosome to maintain correct levels of genetic expression between the sexes, a process termed sex-specific dosage compensation (SSDC). The heteromorphic sex chromosomes (ZZ or ZW) of birds, reptiles and fish do not exhibit sex specific dosage compensation, rather dosage compensation occurs via a distinct mechanism known as network dosage compensation (NDC) (Arnold *et al*., 2008). NDC operates on all chromosomes, and is associated with the coordinate expression of a network of associated genes (Arnold *et al*., 2008).

Like the evolution of sexual differentiation, the mechanisms of dosage compensation between organisms are diverse. Female humans (*Homo sapiens*) and female fruit flies (*D. melanogaster*) have two X chromosomes, whereas the males have an X and a Y chromosome. The Y chromosome has evolved largely from degenerative processes, is gene-poor, and heterochromatic. The Y chromosome plays no role is sex determination but is involved in fertility (Larsson and Meller, 2006). The Y chromosome of the nematode worm *Caenorhabditis elegans* has been lost altogether. The hermaphrodites exhibit a XX genotype, whereas the males are XO (Lucchesi *et al*., 2005). All three organisms exhibit different dosage compensation mechanisms though they achieve the same goal, equalization of X-linked gene product between the two sexes. *D. melanogaster* males overcome their haploinsufficiency (requirement for both alleles of a gene for fitness) by transcriptionally upregulating expression of the single X chromosome roughly two-fold to equal that of the diploid females (Lucchesi, 1973). No universal two-fold upregulation exists, rather individual genes are compensated to a level that ensures fitness (Hamada *et al*., 2005; Straub *et al*., 2005). In humans the females inactivate large regions of one of their X chromosomes, forming a Barr body, to compensate for the males single X chromosome (Lyon, 1961). In *C. elegans* the hermaphrodite downregulates both of its

X chromosomes to compensate for the males single X chromosome (Meyer and Casson, 1986). The DC mechanism of humans and *C. elegans* results in males and females being effectively monosomic for the X chromosome relative to the diploid autosomes and therefore haploinsufficient. Global gene expression profiles indicate that humans and *C. elegans* compensate for this independently by transcriptionally upregulating the X chromosome expression roughly two-fold (Gupta *et al*., 2006; Nguyen and Disteche, 2006) (Figure 1.1). The evolution of the respective DC mechanisms is distinct but they share an underlying paradigm, the modulation of transcription via the formation of transcriptionally permissive or repressive chromatin states. DC is therefore an epigenetic process, a process that alters an organism's genetic expression profile without altering the genetic code of that organism.

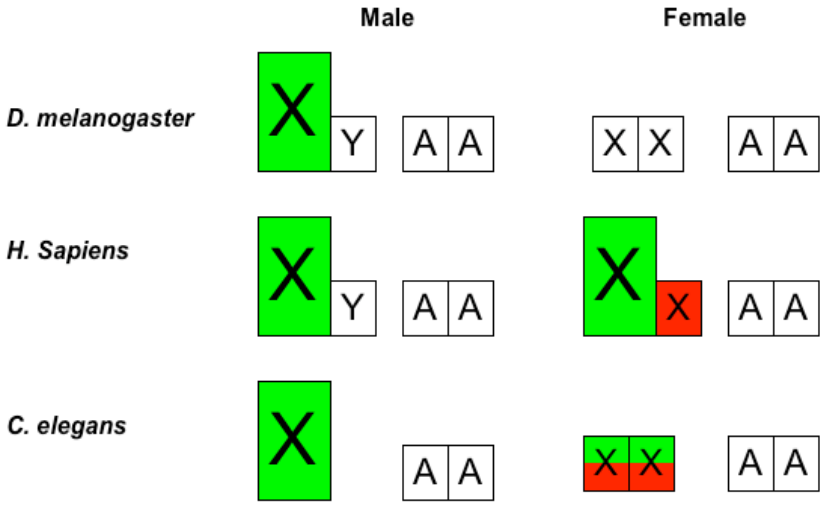


Figure 1.1: Dosage compensation mechanisms. *D. melanogaster* males transcriptionally upregulate their single X chromosome. Female *H. sapiens* inactivate one of their X chromosomes. *C. elegans* hermaphrodites transcriptionally downregulate both X chromosomes. X chromosome expression in *H. sapiens* and *C. elegans* is transcriptionally upregulated to equal that of the diploid autosomes. Activation is represented by green, inactivation by red. Figure reproduced from (Straub and Becker, 2007).

1.3 Sex Determination and Dosage Compensation in *D. melanogaster*

The commitment to a sexual phenotype in any organism begins with a primary signal, which can be environmental or genetic. The primary signal, which differs between sexes initiates a signaling cascade that alters genetic expression in a sex-dependent manner leading to a sexual phenotype. In *D. melanogaster* the sex determination primary signal is the X chromosome to autosome ratio (X:A). Females have a X:A ratio of 1 (2X:2A) whereas males have a ratio of 0.5 (X:2A).

The X:A ratio is *read* by the basic helix-loop-helix protein product of the X-linked *sisterlessB* (*sisB*) gene. In females only, zygotically expressed *sisB* and maternal *daughterless (da)* proteins form heteromeric complexes that bind to the *early* promoter of the *sex-lethal* (*sxl*) gene and control its all-or-nothing expression (Yang *et al*., 2001). *Sxl* encodes an RNA binding protein that functions in sex-specific alternative splicing pathways. The female X:A ratio of 1 is transduced into the transcriptional activation of *early sxl*. The male X:A ratio of 0.5 is transduced into the transcriptional repression of *early sxl*. *Sxl* and its downstream cascade control sexual differentiation of the soma, germline development, and dosage compensation. This occurs early on in embryogenesis around the blastoderm stage and is irreversible. After the blastoderm stage *Sxl* controls its own expression via a positive autoregulatory loop involving an alternative splicing pathway. After the blastoderm stage *sxl* is activated by the *late* promoter and is expressed in both sexes. The absence of *early Sxl* protein in males results in the inclusion of exon L3 in the *sxl* mRNA, which contains a stop codon. *Late Sxl* protein in males is therefore truncated and nonfunctional. In females *early Sxl* protein is involved in the alternative splicing of late *sxl* mRNA, splicing out exon L3. This produces a transcript with the complete open reading frame, which is translated to produce active *Sxl* protein (Cline, 1993).

The *transformer* (*tra*) gene is downstream of *sxl* and is expressed in both males and females. Intron 1 of *tra* has a proximal and a distal 3' splice site. In the absence of *Sxl* protein the proximal 3' splice site is used, which introduces a stop codon. The inclusion of this stop codon produces a truncated, nonfunctional *Tra* protein. In females the presence of *Sxl* protein dictates the use of the distal 3' splice site. The stop codon is no longer introduced and active *Tra* protein is produced. *Tra* forms a heterodimer with *Transformer2* (*Tra2*) protein, which controls the sex-specific splicing of the *doublesex* (*dsx*) and fruitless (*fru*) genes (Bopp *et al*., 1999). The female splice-specific products *FruF* and *DsxF* control the expression of genes

necessary for female differentiation and behavior. In males the absence of *Tra* results in the male splice-specific products *FruM* and *DsxM*, which control the expression of genes necessary for male differentiation and behavior (Figure 1.2).

Sxl is the key determinant in sexual determination and dosage compensation, controlling each via a sex-specific alternative-splicing pathway. Dosage compensation in *D. melanogaster* involves the transcriptional upregulation of the male single X chromosome approximately two-fold. This upregulation is achieved via an epigenetic process that alters the chromatin structure, presumably increasing the accessibility of transcriptional factors or alternatively facilitating the

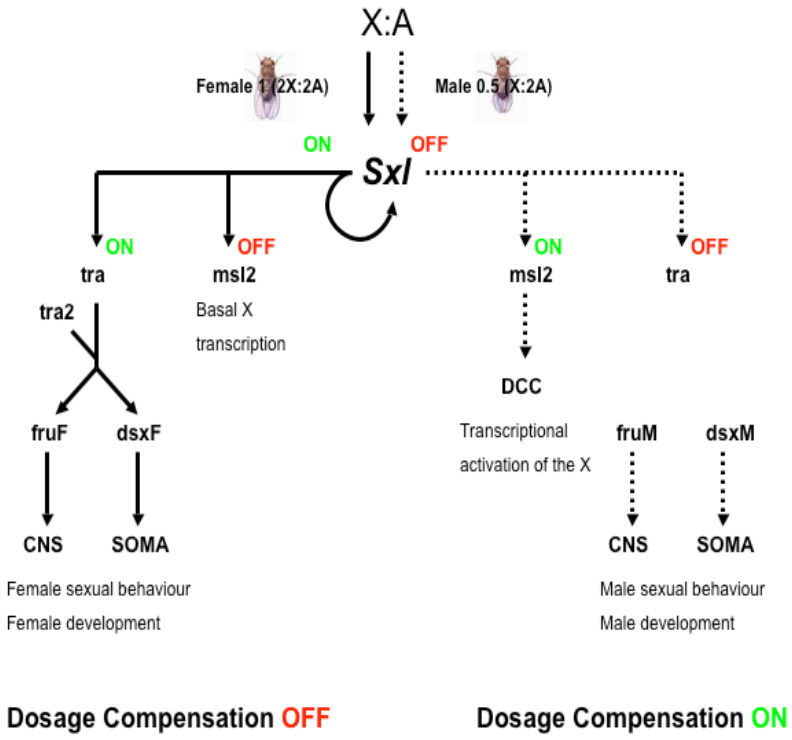


Figure 1.2: Sexual determination and dosage compensation in *D. melanogaster*. At the blastoderm stage of embryogenesis the X:A ratio dictates the all-or-nothing expression of *Sxl*. A X:A of 1 activates *Sxl* expression leading to the development of a female phenotype and inactive dosage compensation. A X:A ratio of 0.5 represses *Sxl* expression leading to a male phenotype and active dosage compensation. The solid line represents the female pathway, the dotted line the male pathway. Based on Figure 1 and Figure 2 from (Penalva and Sanchez, 2003).

elongation process. This altered, decondensed chromatin structure is associated with painting of the male X chromosome with a non-translated RNA and sex-specific global acetylation of lysine 16 of histone 4 (H4K16Ac). Associated with the male specific X chromosome modifications are five male-specific lethal (*msl*) genes: *msl*1, *msl*2, *msl*3, males-absent-on-the-first (*mof*), and maleless (*mle*), and two non-coding RNA molecules: RNA-on-the-X 1 (roX1) and roX2 (Stuckenholz *et al*., 1999). The protein products of these genes form a protein-RNA complex, the dosage compensation complex (DCC) that associates specifically with the male X chromosome. The role that the complex plays in dosage compensation in males will be discussed in greater detail in section 1.5. The *msl* and *roX* genes are transcribed in both males and females however *Sxl* turns off dosage compensation in females by controlling *msl2* expression. *Sxl* represses *msl2* mRNA expression in females by binding to a poly(U) stretch in the 5'UTR, preventing splicing of the first intron. At a later stage, binding of *Sxl* to the poly(U) stretches at the 5' and 3' UTRs prevents translation of the *msl2* mRNA (Bashaw and Baker, 1995). The absence of MSL2 protein in females prevents formation of a functional complex as MSL2 is integral to complex formation and hence dosage compensation does not occur. If the *msl2* cDNA is ectopically expressed from a plasmid a functional DCC forms in females (Chang and Kuroda, 1998).

1.4 Chromatin Structure

In *D. melanogaster*, dosage compensation involves the modulation of transcription via the altering of the local chromatin structure. Chromatin is a hierarchical structure of histone proteins that compacts the meters of DNA found in a eukaryotic nucleus. The functional repeating unit of chromatin is the nucleosome core particle (NCP). The *D. melanogaster* NCP comprises ~146bp of DNA wrapped around an octamer comprised of two H2A-H2B heterodimers and a (H3-H4) tetramer (Clapier *et al*., 2008). The ~146bp of DNA form 1.65 left-handed superhelical turns around the histone octamer. In the yeast NCP structure histone N-terminal tails are relatively disordered and solvent exposed, protruding out from the nucleosome core (White *et al*., 2001) (Figure 1.3).

The nucleosomes are spaced along the linear DNA in a beads-on-a-string array (11nm fiber). The nucleosomes are separated by short stretches of unbound DNA, typically 10-50 bp long (Schwab *et al*., 2008). This string of nucleosomes can compact to from a higher ordered

30nm fiber, in which histone H1 contacts neighboring nucleosomes. The exact structure of the 30nm fiber is a contentious issue. Two major models exist: the one-start solenoidal helix, which is a linear array of coiled nucleosomes with a bent DNA linker between adjacent nucleosomes and the two-start helix where nucleosomes zig-zag to form a ribbon that supercoils, with a straight DNA linker (Dorigo *et al*., 2004; Woodcock *et al*., 1984) (Figure 1.4). The crystal structure of a tetranucleosome exhibits adjacent nucleosomes linked by straight DNA, supporting the two-start hypothesis (Schalch *et al*., 2005; Wu *et al*., 2007). The hierarchical structure continues until the metaphase chromosome is achieved.

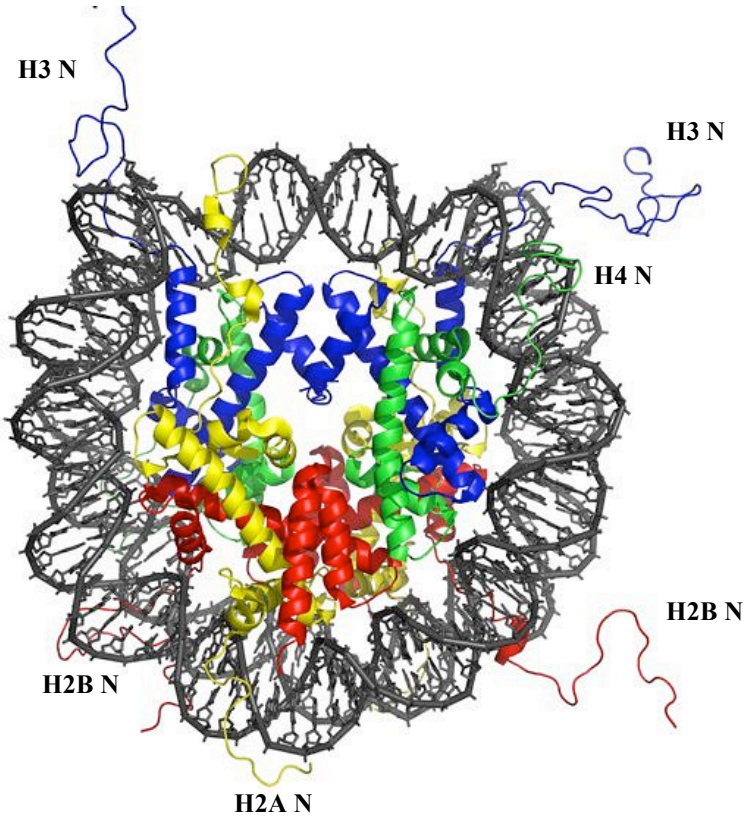


Figure 1.3: Structure of the yeast *Saccharomyces cerevisiae* nucleosome core particle PDB 1id3 (White *et al*., 2001). Histone H2A is shown in yellow, H2B in red, H3 in blue and H4 in green. DNA is shown in turquoise. Figure taken from http://en.wikipedia.org/wiki/Nucleosome, with permission from Richard Wheeler (Zephyris) at wikipedia.org).

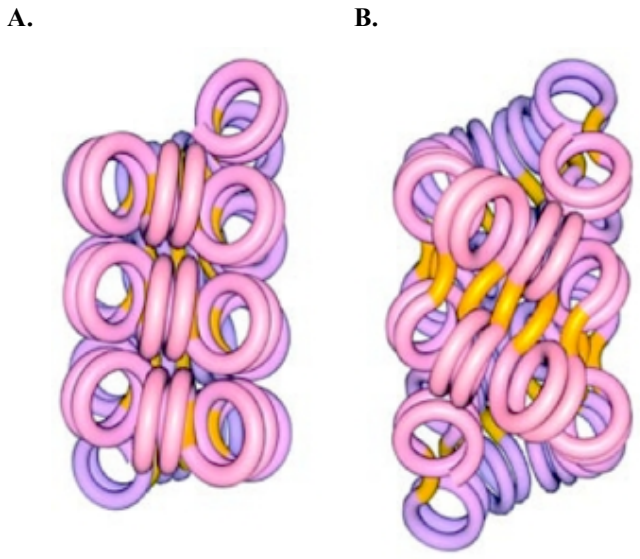


Figure 1.4: Models of the 30 nm chromatin fiber. Two major models exist: the one-start solenoidal helix, which is a linear array of coiled nucleosomes (**A**), and the two-start helix where nucleosomes zig-zag to form a ribbon that supercoils (**B**) (Dorigo *et al*., 2004).

Chromatin structure is highly dynamic and its local structure can impede or promote all essential DNA processes such as recombination, DNA repair, replication and transcription. The two major chromatin states found within the cell nucleus are heterochromatin and euchromatin. Heterochromatin's condensed structure represses transcriptional activity. Heterochromatic regions include repetitive satellite sequences, centromeres, telomeres, and the inactive X chromosome (Barr body) of female mammals (Lohe *et al*., 1993). Heterochromatin can be either constitutive or facultative. Constitutive heterochromatin is defined as regions of heterochromatin identical in all cells of a species. The Y chromosome is largely constitutive heterochromatin. Facultative heterochromatic is defined as regions of heterochromatin that differ between cells of an organism. X chromosome inactivation in female mammals involves the formation of facultative heterochromatin (Craig, 2005). Unlike euchromatin, heterochromatin is only found in eukaryotes, not prokaryotes. Euchromatin has a diffuse, decondensed structure, which is transcriptionally permissive. Euchromatic regions are generally gene-rich and actively transcribed. Eukaryotic cells can control replication and the differential expression of a gene by transforming from heterochromatin to euchromatin. Elaborate mechanisms have evolved to modulate chromatin structure. Modulation of chromatin structure to facilitate either a transcriptionally repressive or permissive state hinges around changes in

nucleosome histone composition , ATP-dependent chromatin remodeling, and the reversible post-translational modification of the histone N-terminal tails, which protrude out from the NCP. Modulation of chromatin structure is quintessential to dosage compensation.

1.4.1 Histone Variants

The canonical nucleosome is composed of histones H2A, H2B, H3, and H4. Histone H1 acts as a linker between nucleosomes. A number of variants are known for all histones, except H4. The function of the individual variants is diverse, ranging from transcriptional activation, recombination and repression to DNA repair. Histone variant incorporation modulates chromatin structure via altering intra- and inter-nucleosomal contacts and potential new targets for unique post-translational modifications. The role of histone post-translational modification is discussed in-depth in a later section (1.5.1). Unlike the canonical histones, which are encoded by multiple genes, a single gene encodes the various histone variants. Variant histone genes fall into one of three classes: replication dependent (RD), replication and cell-cycle independent (RI), and tissue specific (TS) based upon their expression. The dependence upon replication refers to whether or not the histone is deposited during DNA replication, behind the progressing replication fork. The nucleosome code theory defines the hypothesis that functionally distinct homotypic nucleosomes are formed by the specific incorporation of a distinct histone variant. The incorporation of different variants confers distinct functions (Bernstein and Hake, 2006). See Table 1.1 for a synopsis of histone variants, their class and function.

Histone H1 expression is highly specific to tissue, species, and developmental stage. The H1 family is large and members are often redundant, functioning in chromatin condensation via forming contacts with neighboring nucleosomes. Binding of H1 and its variants to neighboring nucleosomes and the *linker* DNA between nucleosomes is transient (Lever *et al*., 2000). Deletion studies demonstrated that cells can exist without, or with drastically reduced levels of H1 and that deletion of one variant is compensated for by the increased expression of another variant (Fan *et al*., 2003; Takami *et al*., 2000; Wierzbicki and Jerzmanowski, 2005).

Four major H2A variants are known H2A.X, H2A.Z, macroH2A, and H2ABbd (Barr body deficient). While H2A.X is the main form in the yeast *S. cerevisiae*, in mammals this histone functions as a DNA damage sensor. Phosphorylated H2A.X, denoted γ-H2A.X, is

enriched at double-strand break points (Redon *et al*., 2002). H2A.Z functions in gene silencing in mammals (Rangasamy *et al*., 2003). The *D. melanogaster* H2A.Z homologue, H2AvD plays an important role in, but is not specific to dosage compensation. H2AvD, is involved in gene activation and silencing (Zlatanova and Thakar, 2008). Ubiquitination of its C-terminal tail or acetylation of its N-terminal tail differentiates between silencing and activation, respectively (Sarcinella *et al*., 2007). In terms of dosage compensation in *D. melanogaster*, the histone variant H2AvD is incorporated into, and helps maintain euchromatic regions via acetylation of its N-terminal tail (Tanabe *et al*., 2008). In female mammals incorporation of the H2A variants macroH2A1.1 and macroH2A1.1, splice variants of H2A1 is essential to X inactivation, and ultimately dosage compensation (Chow and Brown, 2003). MacroH2A is vertebrate-specific and is very large. It is incorporated into chromatin of the inactive mammalian female X chromosome (Barr body), functioning in transcriptional repression (Costanzi and Pehrson, 1998). As its name indicates H2A.Bbd is Barr body deficient. It is the most distantly related H2A variant and associates with regions of hyperacetylated H4, suggesting a role in transcriptional activation (Chadwick and Willard, 2001). Three H2B variants are known, all of which are testis/sperm specific. The function of human-sperm-specific H2B (spH2B), human testis/sperm-specific H2B (hTSH2B), and H2BFWT are not well characterized but nucleosomes containing these variants are less stable and may confer resistance to apoptosis (Cheung *et al*., 2003; Li *et al*., 2005a).

To date five histone H3 variants have been described in mammals, H3.1, H3.2, H3.3, H3.1t and CENP-A. It is suggested that H3.1, H3.2, and H3.3 are incorporated into constitutive heterochromatin, facultative heterochromatin, and euchromatin respectively (Hake and Allis, 2006). In *D. melanogaster* the histone H3.3 variant plays an important role in, but is not specific to dosage compensation. Absent from heterochromatin, H3.3 is enriched in euchromatin of the male X chromosome, throughout all active genes (Wirbelauer *et al*., 2005). H3.3 deposition is linked to RNA polymerase II activity and correlates with dimethylated H3 K4, an epigenetic mark associated with transcriptionally active chromatin (Mito *et al*., 2005). H3.1t is a testis specific variant (Witt *et al*., 1996). The H3 variant CENP-A is a critical centromeric protein, localized exclusively to the active kinetochore (Sullivan, 2001).

While the nature of DNA-packaging into discreet nucleosomes and the organization of said nucleosomes into higher ordered structures is highly conserved, the authors of an

Table 1.1: H2A, H2B, and H3 variants. Synopsis of histone H2A, H2B, and H3 variant expression, chromosomal localization, replication dependence (RD or RI) and function. In *D. melanogaster* (*D. m)*, the histone variant H2AvD is involved in dosage compensation via maintain euchromatic regions via acetylation of its N-terminal tail (Tanabe *et al*., 2008).

Histone variant	Expression	Localization	Function
H2AX	RI	All	DNA repair
H2A.Z	RI	All	Activation and silencing
H2Av (*D. m)*	RI	All	Activation and silencing (Dosage Compensation)
H2A.Bdb	unknown	not Inactive X	Activation
MacroH2A	unknown	Inactive X	Inactivation
spH2B	unknown	Telomeres	Unknown
hTSH2B	unknown	unknown	Unknown
H2BFWT	unknown	Telomeres	Unknown
H3.1	RD	Constitutive H	Inactivation
H3.2	RD	Facultative H	Inactivation
H3.3	RI	Euchromatin	Activation (Dosage Compensation)
H3.1t	RD	unknown	Unknown
CENP-A	RI	Centromeres	Segregation

exhaustive review of histone variants suggest that the site-specific incorporation of histone variants exists to maintain dynamic regulation of chromatin structure through altered nucleosome structure and distinct post-translational modifications specific to the respective histone variant (Bernstein and Hake, 2006).

1.5 Dosage Compensation in *D. melanogaster*

In *D. melanogaster*, associated with the two-fold up regulation of transcription is the male-specific lethal (MSL) complex, a RNA-protein complex comprised of at least five known proteins; MSL1, MSL2, MSL3, MOF, and MLE and two non-translated RNA molecules; roX1 (RNA on the X chromosome) and roX2 that directly modulates the chromatin structure of the

male X chromosome (Lucchesi, 1998; Straub and Becker, 2007). MOF and MLE have enzymatic activity. MLE is a DExH box /DNA helicase with ATPase activity (Lee *et al*., 1997). MOF is a MYST (MOZ, YBF2/SAS3, SAS2, Tip60) family histone acetyl transferase (HAT) (Akhtar and Becker, 2000). MYST family HATs are characterized by a unique C_2HC zinc finger and a N-terminal chromodomain. The MYST family is one of three distinct groups of N^{ε}-lysine acetyltransferases, the other groups are Gcn5/PCAF and p300/CBP (Yang, 2004). A model for the assembly of the dosage compensation complex on the male X chromosome is shown in Figure 1.5.

1.5.1 Histone Post-translational Modification and Dosage Compensation

Histones are subject to many types of reversible post-translational modifications, including methylation (mono-, di-, or tri-), phosphorylation, ubiquitination, acetylation, sumoylation, ADP-ribosylation, and biotinylation (Peterson and Laniel, 2004). These modifications are typically localized to lysine (K), arginine (R), threonine (T), serine (S), glutamic acid (E), histidine (H), and tyrosine (Y) residues in the histone N-terminal tail, which protrude from the nucleosome core particle. Histone C-terminal tails are also capable of being post-translationally modified, but they are not as common or as intrinsic to epigenetic regulation as post-translational modifications of the histone N-terminal tails (Fingerman *et al*., 2007). All four histone N-terminal tails are highly post-translationally modified. It is likely that the conformation of the histone N-terminal tails is highly flexible, and that contacts between the histone tails and DNA is transient. Histone post-translational modification is thought to function through directly altering intra- and inter-nucleosomal interactions (*cis* acting) and/or by creating docking sites for effecter molecules (*trans* acting). It is proposed that the distinct modifications of one or more histone tails can act sequentially or in concert to initiate a distinct chromatin modulating cascade, a theory defined as the histone code hypothesis (Strahl and Allis, 2000).

Histone acetylation is a post-translational modification intrinsic to transcriptional regulation in eukaryotes. Histone acetylation is predominantly associated with upregulation, however a repressive function has also been demonstrated in *D. melanogaster* and yeast (Grienenberger *et al*., 2002). Histone acetylation alters nucleosome-DNA and nucleosome-nucleosome interactions (*cis*-acting) (Hong *et al*., 1993; Luger *et al*., 1997) and acetylated

histones are directly bound by bromodomains (*trans*-acting) (Dhalluin *et al*., 1999). The incorporation of H4 homogenously acetylated at K16 into nucleosomal arrays prevents formation of the 30 nm fiber and inter-fiber contacts *in vivo*, suggesting that the modification may in part be *cis*-acting (Shogren-Knaak *et al*., 2006). H4K16 has been proposed to interact with an acidic residue rich patch on the surface of the H2A-H2B dimer of a neighboring nucleosome (Luger *et al*., 1997). Acetylation of H4K16 would therefore presumably disrupt this interaction leading to a decondensed chromatin structure (Luger *et al*., 1997). Alternatively, H4K16Ac recognition by a bromodomain could lead to the recruitment of a complex that catalyses chromatin decondensation. Bromodomains are one of a small family of distinct histone binding modules found in chromatin binding proteins that recognize specific

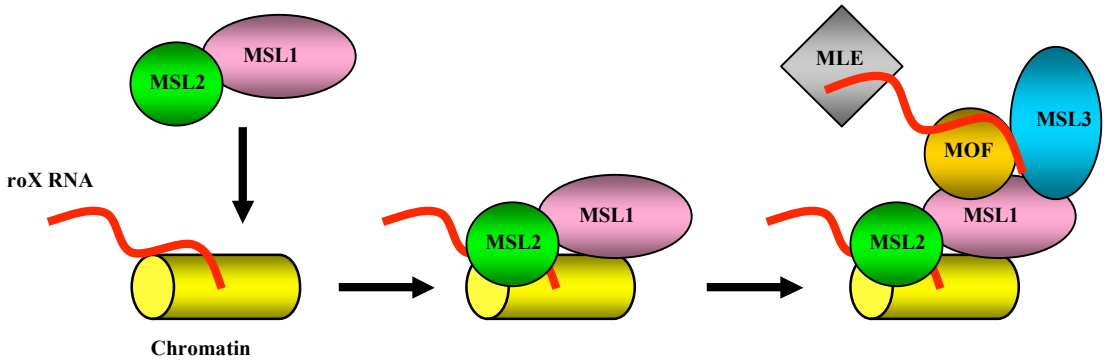


Figure 1.5: Model for the assembly of the *D. melanogaster* dosage compensation complex on the male X chromosome. MSL1-MSL2 forms the core of the complex and can autonomously interact with the X chromosome. MOF and MSL3 interact with MSL1. MLE is likely to transiently interact with the roX RNA. All of the components are essential for a functional dosage compensation complex (Kelley *et al*., 1999).

histone post-translational modifications (Taverna *et al*., 2007) (Table 1.2). Hence both *cis*-acting and *trans*-acting mechanisms are plausible options for nucleosome decondensation.

Acetylation of H4 K16 (H4K16Ac) is a post-translational modification enriched upon the male single X chromosome. It is catalyzed by MOF and is essential to dosage compensation in *D. melanogaster* (Akhtar and Becker, 2000; Smith *et al*., 2000). In the absence of MOF the H4K16Ac of the male X chromosome is rapidly lost (Akhtar and Becker, 2000). H4K16Ac is deacetylated by the histone deacetylase Rpd3 (Buscaino *et al*., 2003). The counteracting activities of MOF and Rpd3 are likely to regulate dosage compensation.

Histone acetyltransferases, such as the MYST family member MOF catalyze the transfer of an acetyl group from acetyl-coenzyme A (AcCoA) to the ζ-nitrogen of a lysine's amino group. The crystal structures of the MYST family yEsa1 (Essential Sas-related acetyltransferase 1) and the Gcn5/PCAF family Gcn5 HAT domains have been reported at high resolution (Trievel *et al*., 1999; Yan *et al*., 2002). Despite belonging to separate HAT families their structures reveal a structural conservation of the central core involved in AcCoA binding and both share a conserved glutamate residue essential to catalysis (Yan *et al*., 2002). Despite this observed structural conservation, Gcn5 and Esa1 were proposed to operate by separate mechanisms. Gcn5 catalyses the direct nucleophilic attack of the AcCoA by the deprotonated ζ-nitrogen of lysine (Trievel *et al*., 1999). Esa1 was initially reported to have a ping-pong mechanism, transfer of the AcCoA acetyl group to the deprotonated ζ-nitrogen of lysine proceeding via an acetyl-Cys residue intermediate (Yan *et al*., 2002). Recent studies using recombinant Esa1 incorporated into the piccolo NuA4 complex suggests that its mechanism actually proceeds via direct nucleophilic attack (Berndsen *et al*., 2007). The piccolo NuA4 (picNuA4) complex is a smaller NuA4 complex, containing only Esa1, Epl1, and Yng2 (Boudreault *et al*., 2003). MOF, a MYST family member is likely to have a similar mechanism based upon conservation of essential residues.

Integration of MOF into the dosage compensation complex confers a male X chromosome specific H4K16Ac distribution, which leads to a decondensed chromatin structure and facilitates its two-fold transcriptional upregulation (Turner *et al*., 1992). The decondensed chromatin structure of the male X chromosome likely facilitates transcriptional upregulation through permitting greater accessibility of RNA polymerase and its associated factors. The yeast *S. cerevisiae* HATs Gcn5 (part of the SAGA complex) and Esa1 (part of the NuA4

complex) specifically acetylate promoter regions (Bhaumik and Green, 2001; Smith *et al*., 1998). Acetylation at the 5' promoter region is proposed to promote transcription initiation by creating a nucleosome free region (Allard *et al*., 1999; Kuo *et al*., 1998), recruitment of TATA-binding protein (TBP) and RNA polymerase II (polII) (Qiu *et al*., 2004) and coactivators (Govind *et al*., 2005). The enriched 3' distribution of H4K16Ac by MOF in the dosage compensation complex is proposed to facilitate transcriptional elongation on the male X chromosome rather than transcriptional initiation (Smith *et al*., 2001).

Table 1.2: Chromatin binding domains. Synopsis of histone post-translational modifications (methylation [Me], phosphorylation [ph], acetylation [Ac]) and the conserved domains that recognize them. Table is based upon Figure 1 from (Taverna *et al*., 2007).

Domain	PTM	Example
Bromodomain	KAc	Gcn5
Tudor Royal Family		
Chromodomain	H3K9Me2/3, H3K27Me2/3	HP1
Double chromodomain	H3K4Me1/2/3	CHD1
Chromo barrel	H3K36Me2/3	Eaf3
Tudor	RMe2	
Tandem Tudor	H3K4Me3, H4K20Me3, H4K20Me1/2	JMJD2A
MBT	H4K20Me1/2, H1K26Me1/2, H3K4Me1, H3K9Me1/2	L3MBTL1
PHD finger	H3K4Me3, H3K4Me, H3K9Me3	BPTF
WD40	H3R2/K4Me2	WDR5
14-3-3	H3S10ph, H3S28ph	14-3-3ξ
BRCT	H2AX-S139ph	BRCA1

1.5.2 ATP-dependent Nucleosome Remodeling and Dosage Compensation

The regulators and enzymes of transcription, recombination, replication and repair require access to their respective DNA targets, which a nucleosome would otherwise block. A repressive local chromatin structure is altered to permit these processes by nucleosome remodeling complexes. Nucleosome remodeling complexes catalyze the ATP-dependent sliding of nucleosomes to neighboring DNA sequences. Nucleosome remodelers within all eukaryotes fall into one of five families: SWI/SNF, ISWI, NURD/Mi2/CHD, and INO80/SWR1 (Becker and Horz, 2002). The molecular engines of these diverse complexes are homologous ATPases. Homologues are differentiated based upon sequence variability outside of their ATPase domain. The first class of ATPase, Swi2/Snf2, was identified in *S. cerevisiae* in a screen that identified the *switch* of mating type and *sucrose-nonfermenting* mutations and is part of the 11-subunit SWI/SNF complex (Sudarsanam and Winston, 2000; Vignali *et al*., 2000). The Brahma (BRM) ATPase, a SWI/SNF homologue in *D. melanogaster* is part of the BRM complex (Elfring *et al*., 1994). BRM is involved in RNA polymerase II (Pol II) regulation, elongation and development (Armstrong *et al*., 2002; Srinivasan *et al*., 2005; Zraly *et al*., 2004). The ATPase ISWI (imitation switch) was later identified in *D. melanogaster* based upon homology with BRM. ISWI ATPases are characterized by two SANT-like DNA-binding domains (Aasland *et al*., 1996). ISWI is present in three complexes, NURF (nucleosome remodeling factor), CHRAC (chromatin accessibility complex), and ACF (ATP-utilizing chromatin assembly and remodeling factor) (Langst and Becker, 2001). These complexes function in the formation and disruption of higher ordered chromatin structures, which are repressive to fundamental DNA processes. The antagonistic effects of these complexes depends whether they catalyze a succession of nucleosomes into an ordered array (Varga-Weisz *et al*., 1997), which would facilitate the formation of a higher ordered chromatin structure or catalyze the disruption of ordered nucleosomal arrays (Tsukiyama and Wu, 1995).

ATP-dependent nucleosome remodeling complexes play a key role in dosage compensation in *D. melanogaster* by fine-tuning the two-fold transcriptional activation of the male X chromosome. ATP-dependent nucleosome remodeling complexes that disrupt ordered nucleosomal arrays promote euchromatin formation and would therefore be complimentary to the function of the DCC. Complexes that catalyze the formation of ordered nucleosomal arrays promote heterochromatin formation, which would be antagonistic. The function of the ISWI

ATPase appears to be directly antagonistic to MOF, catalyzing ordered nucleosomal arrays, which promotes a repressive chromatin structure. Homozygous *ISWI* null mutant male larvae exhibit a highly decondensed X chromosome. The autosomes and female X chromosomes exhibit a wild-type phenotype. This decondensed male phenotype is not observed in larvae with homozygous *ISWI* and *mle* mutations (Corona *et al*., 2002). MLE is essential to the function of the DCC (Gu *et al*., 2000), therefore the function of ISWI and the DCC is antagonistic. The N-terminal tail of histone H4 alone is pivotal to the interplay between these complexes in dosage compensation. When the H4 N-terminal tail is deleted ISWI loses its ATPase activity and NURF and CHRAC subsequently lose their ability to catalyze the nucleosome sliding. (Clapier *et al*., 2001; Hamiche *et al*., 2001). The crucial component of the H4 histone N-terminal tail for ISWI activity is the stretch of residues $K_{16}R_{17}H_{18}R_{19}$ (Clapier *et al*., 2002). It is the DCC dependent H4K16 acetylation of this stretch of residues that specifically antagonizes ISWI function. Acetylation of H4K16 inhibits ISWI ATPase activity and nucleosomal sliding (Corona *et al*., 2002).

The non-translated RNA roX1 and roX2 are integral components of the DCC in *D. melanogaster* and function in complex assembly and targeting to the males single X chromosome (Kelley *et al*., 1999; Peterson and Laniel, 2004). Like H4K16 acetylation, the roX RNAs are essential to the antagonistic interplay between ATP-dependent nucleosomal remodelers and the DCC in *D. melanogaster*. NURF is a four subunit complex (NURF301, NURF55, NURF38, and ISWI) that functions as an activator and repressor of transcriptional activity (Badenhorst *et al*., 2005; Xiao *et al*., 2001). Like in the homozygous *ISWI* null mutant, the male X chromosome exhibits a highly decondensed structure in *NURF301* mutant males. This phenotype is not suppressed if either *roX1* or *roX2* is concomitantly deleted. Interestingly, unlike H4K16 acetylation, which suppresses the action of ISWI, NURF represses the action of the DCC by repressing transcription of the *roX* genes (Bai *et al*., 2007).

ATP-dependent nucleosomal remodeling complexes also have the ability to generate positive superhelical torsion in DNA, which further propagates the formation of repressive higher ordered chromatin. It is therefore not surprising that DNA supercoiling factor (SCF), a protein that introduces negative supercoils, which would be antagonistic to the action of ATP-dependent nucleosome remodelers is associated with the action of the DCC. SCF interacts

directly with the DCC and when overexpressed causes the male X chromosome to be highly decondensed. This phenotype is suppressed by the *ISWI* mutation (Furuhashi *et al*., 2006).

The *D. melanogaster* GAGA factor, encoded by *Trithorax-like* (Trl) functions in the formation of nucleosome-free regions of chromatin and is implicated in DC. *Trl* hypomorphs (a mutant gene with weaker effect than the corresponding wild-type gene) show reduced male viability and redistribution of the DCC to autosomes (Greenberg *et al*., 2004).

1.6 Dosage Compensation Complex: Structure and Localization

Dosage compensation in *D. melanogaster* involves the two-fold transcriptional upregulation of the males single X chromosome, a finely tuned epigenetic process that involves the coordinated post-translational modification and remodeling of the chromatin structure. Central to the modulation of the chromatin structure is the ribonucleoprotein DCC, which *paints* the male X chromosome and is responsible for global H4K16 acetylation (Akhtar and Becker, 2000). A functional DCC comprises MSL1, MSL2, MSL3, MOF, MLE, roX1 and/or roX2 (Belote and Lucchesi, 1980). The two roX RNAs share very little sequence similarity and appear to be redundant in their function (Meller and Rattner, 2002). Knockout of either subunit of the DCC abolishes DC. MSL1 and MSL2 form the core of the complex, which can bind to approximately thirty to sixty cytologically distinct sites on the male X chromosome in an *msl3*, *mof*, or *mle* knockout (Demakova *et al*., 2003; Kelley *et al*., 1999). MSL1 recognizes these high affinity sites via a conserved basic motif at the N-terminus (1-26). Heterodimerization with MSL2 is essential for recognition (Akhtar and Becker, 2000; Li *et al*., 2005b). MSL1 and MSL2 heterodimerize via a two-stranded, α-helical, coiled-coil structure (Scott *et al*., 2000). Coiled-coil structures are formed by a heptad repeat $(abcdefg)_n$ where *a* and *d* are typically small hydrophobic residues and *e* and *g* are charged amino acids (Cohen and Parry, 1990). The region of MSL2 that is likely to form the coiled-coil (25-40) precedes a RING finger domain, the integrity of which is essential to interaction with MSL1 (Copps *et al*., 1998). The region within MSL1 that forms the coiled-coil structure is at the N-terminus, C-terminal to the nucleosomal binding basic motif (Li *et al*., 2005b). Li *et al*. (2005b) noted that a truncated MSL1 construct containing only the first 265 amino acids (1-265) was able to bind hundreds of sites along the male X chromosome, but was only able to bind the approximately thirty high affinity sites in the absence of endogenous MSL1 (Li *et al*., 2005b). Experimental evidence

demonstrated that this is due to self-association via a glycine-rich motif (26-50), between the basic motif and the coiled-coil motif allowing for the truncated mutant to bind to wild-type complex. Li *et al*. (2005b) suggest that complex formation at these high affinity sites facilitates the cooperative binding of the complex to proximal sites of weaker affinity via self-association of MSL1. Interactions of the N-terminus of MSL1 are summarized in Figure 1.6.

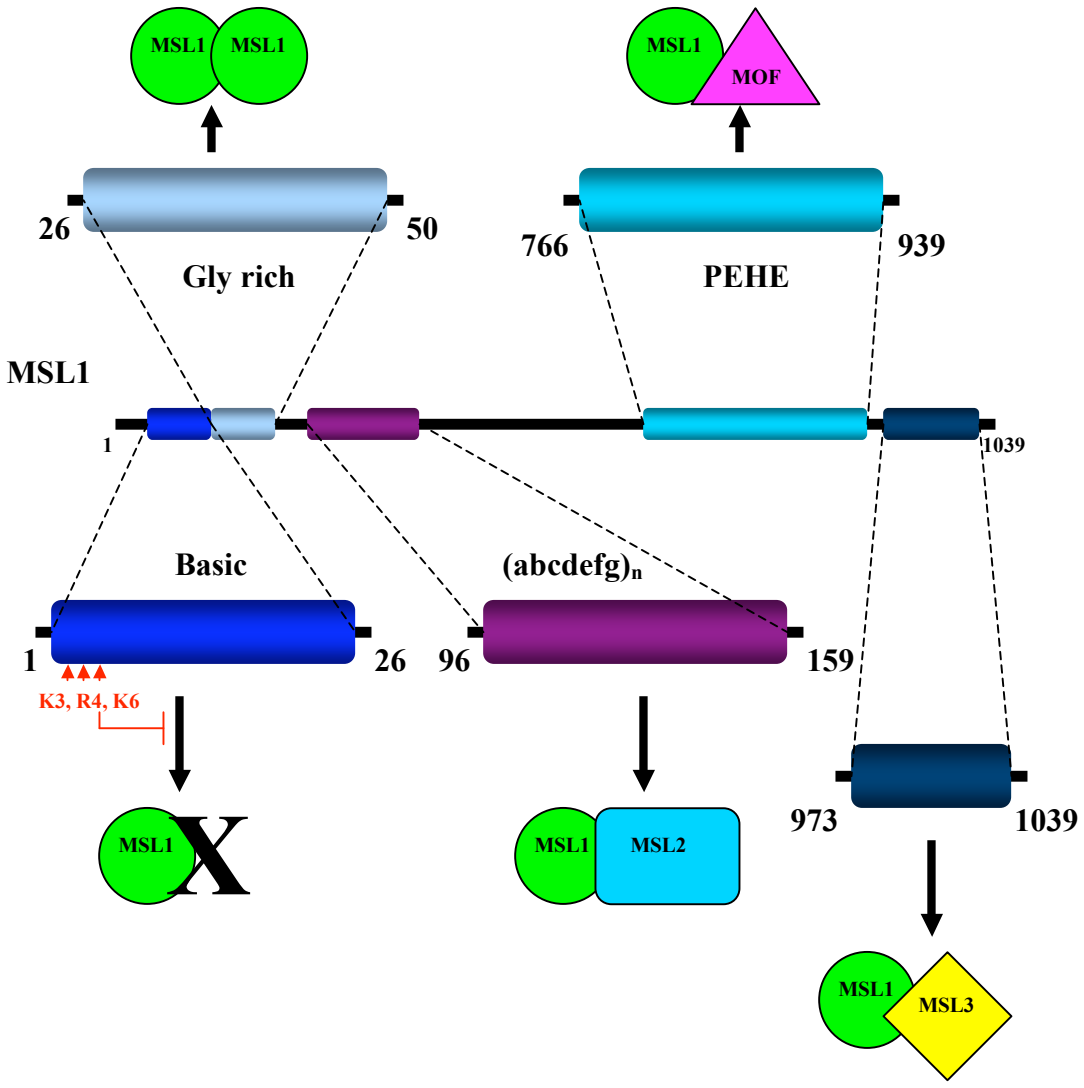


Figure 1.6: Schematic representation of the *D. melanogaster* MSL1 protein. Its N-terminus is involved in binding the male X chromosome (1-26), self association (26-50) and interaction with MSL2 (96-159). The C-terminus is involved in binding MSL3 and MOF. Mutation of K3, R4, and K6 restricts binding of MSL1 to approximately 5 of the high affinity entry sites (Li *et al*., 2005b; Smith *et al*., 2005).

Interestingly, two of the thirty high affinity sites are the *roX1* and *roX2* genes. Despite differential expression roX1 and roX2 appear to be functionally redundant (Meller, 2003; Meller and Rattner, 2002). The concomitant deletion of roX1 and roX2 however causes a dramatic relocalization of the DCC to the autosomes and pericentric heterochromatin (Meller and Rattner, 2002). The roX RNAs therefore play a role in the correct localization of the complex. The key determinant in conferring male X chromosome specificity appears to be the C terminus of MSL2 (Marin, 2003). Binding of roX1 and its incorporation into the complex requires proline-rich and basic motifs in the C-terminus, with the aid of MLE, which can bind roX1 autonomously and weakly interacts directly with the C-terminus of MSL2 (Li *et al*., 2008; Meller *et al*., 2000).

While the N-terminus of MSL1 and MSL2 are sufficient for recognition of pericentromeric heterochromatin and the X chromosome high affinity sites, normal binding of the hundreds of sites along the X chromosome requires proline rich and basic motifs in the C-terminus of MSL2 (Li *et al*., 2008; Meller *et al*., 2000).

MOF, MLE, and MSL3 have been demonstrated to bind RNA (Akhtar *et al*., 2000; Izzo *et al*., 2008; Morales *et al*., 2005; Richter *et al*., 1996). It is therefore likely that they associate with the MSL1/MSL2 core concomitant with, or after roX incorporation. Binding of roX RNA by MOF and MSL3 is not crucial to integration into the DCC (Oh *et al*., 2003). The C terminus of MSL1 is the key determinant involved in the direct binding of MOF and MSL3. MOF binds to a region of the MSL1 C terminus that contains a conserved PEHE domain (766-939) via its C_2HC zinc finger (Marin, 2003; Morales *et al*., 2004). MOFs C_2HC zinc finger is also essential to autonomous nucleosome recognition and its HAT activity (Akhtar and Becker, 2001). The MOF binding site is distinct from the MSL3 binding site (973-1039) (Morales *et al*., 2004). MSL3 binds to the C-terminus of MSL1 (973-1039) via its C –terminal MRG domain (Morales *et al*., 2005). The domain structure and interactions between components of the dosage compensation complex are shown in Figure 1.7.

The likely dual function of MOF has been previously discussed along with the male specific global enrichment of H4K16Ac on the male X chromosome, which is catalyzed by MOF incorporated into the DCC. Specificity of MOF for H4K16 of male X chromosome nucleosomes is conferred synergistically by complex formation with MSL1 and MSL3 (Morales *et al*., 2004). MOF and MSL3 must be bound to MSL1, and MOF and MSL3 must

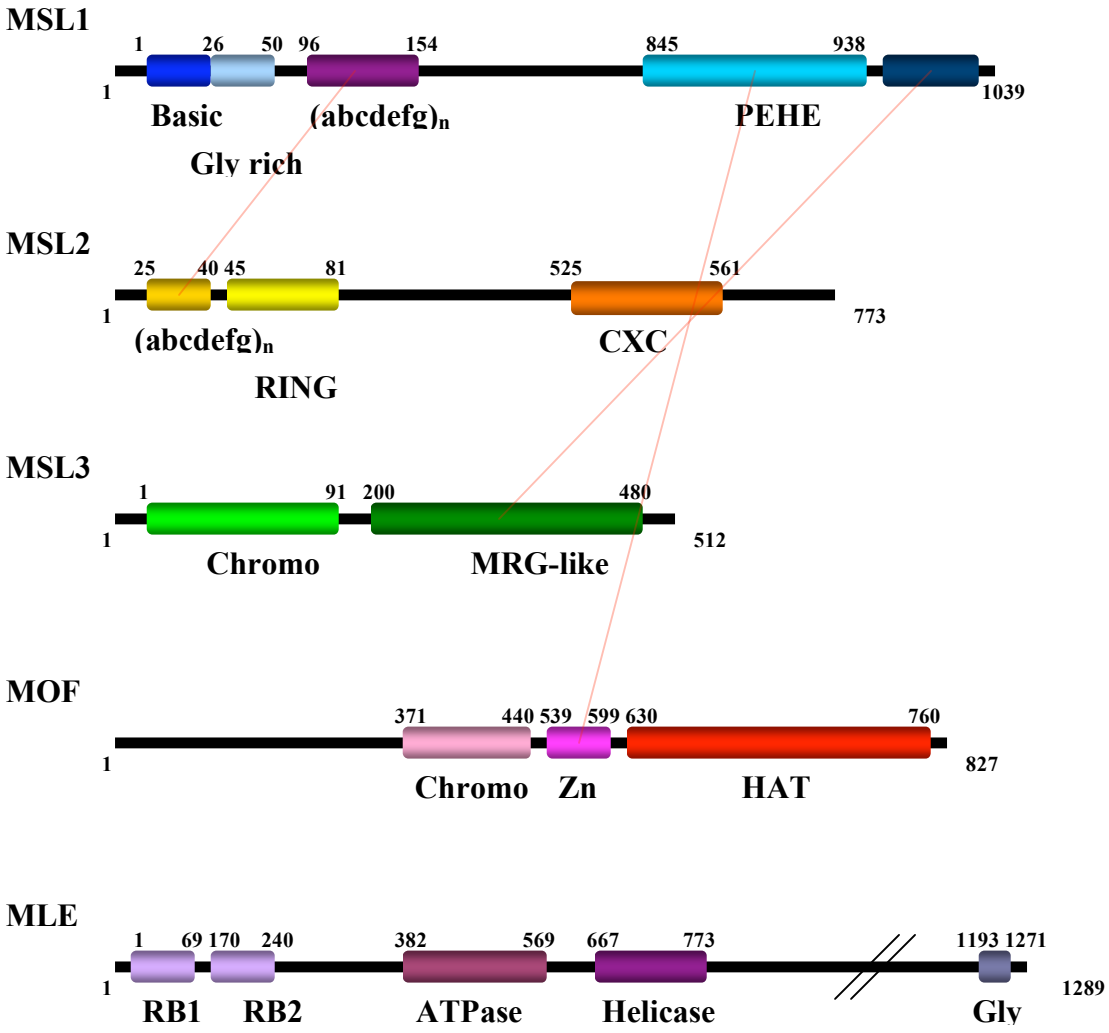


Figure 1.7: Schematic representation of the MSL1, MSL2, MSL3, MOF, and MLE domain structure. Interactions are denoted by a red line.

directly interact for specificity and increase in acetylase activity (Morales *et al*., 2004). Morales *et al*. (2004) also noted that when not in complex MOF autoacetylated and acetylated MSL1 and MSL3. The acetylation state of MSL3 may be important to its incorporation into the DCC. MSL3, specifically acetylated at K116 by MOF no longer localizes to the X chromosome, implying that acetylation prevents its integration into the DCC (Buscaino *et al*., 2003). Note, some skepticism exists as to whether MSL3's incorporation into the DCC is acetylation-dependent as K116 is not conserved in other *Drosophila* homologues (personal communication Dr. S. A. Moore). The findings of Buscaino *et al*., 2003 have also not been corroborated by other laboratories.

Acetylation of MSL3 at K116 prevents X chromosome localization via interrupting the interaction with roX2. MOF-dependent acetylation of MSL3 is counteracted by the histone deacetylase Rpd3, which interacts directly with MSL3 (Buscaino *et al*., 2003). Rpd3 is also likely to be responsible for the deacetylation of H4K16Ac (Akhtar and Becker, 2000). This interplay of two enzymes with opposing activities in the regulation of chromatin post-translational modification and ultimately chromatin structure is not unique. The *S. cerevisiae* NuA4 complex, which contains the MOF HAT paralogue Esa1 has previously been discussed. The complex also contains an MSL3 paralogue, Eaf3 (Esa1 associated factor 3). Eaf3 governs the complex's specificity for acetylating H3 and H4 of nucleosomes at promoter regions over protein-coding regions (Joshi and Struhl, 2005; Morales *et al*., 2005; Reid *et al*., 2004). Eaf3 is found in the NuA4 HAT complex and in the Rpd3S and Rpd3L HDAC complexes (Carrozza *et al*., 2005; Gavin *et al*., 2002; Reid *et al*., 2004). Acetylation of ORF nucleosomes facilitates transcriptional elongation. To prevent intragenic transcription initiation this post-translational modification is removed after the passage of RNApol II via methylation of H3K36 by the histone methyltransferase SET2, which directly interacts with the C-terminal domain of RNApol II via its Set2 Rpb1 interacting (SRI) domain (Kizer *et al*., 2005; Krogan *et al*., 2003). The H3K36Me_3 post-translational modification is then recognized by the Rpd3S complex via the coupled action of the chromodomain of Eaf3 and the PHD domain of Rco1 (Joshi and Struhl, 2005; Li *et al*., 2007a). The coupled recognition of H3K36Me3 explains why the NuA4 complex, which contains Eaf3 but not Rco1 does not recognize H3K36Me_3 modified nucleosomes (Li *et al*., 2007a). Eaf3 therefore plays a crucial role in linking Rpd3 catalyzed histone deacetylation to RNApol II elongation in yeast. It is unclear if MSL3 functions in

distinct HAT and HDAC complexes in *D. melanogaster* but the localization of the DCC and specifically MSL3 have been demonstrated to overlap with the localization of H3K36Me3 and H3K4Me modified nucleosomes (Larschan *et al*., 2007). Methylation of H3K36 by the methyltransferase Hypb in *D. melanogaster* has been demonstrated to recruit the DCC and promote H4K16 acetylation (Bell *et al*., 2008). The highly conserved nature of the interaction and the similar actions of the DCC and NuA4 complex suggests that MSL3 will play a similar role to Eaf3.

Along with the histone deacetylase Rpd3, a histone kinase, JIL1 is also transiently associated with the DCC. JIL1 is an Aurora B homologue that specifically phosphorylates H3S10 and associates with gene-rich interband regions. It is essential to both sexes, however is enriched upon the male X chromosome (Jin *et al*., 1999). JIL1 hypomorphs (a mutant gene with weaker effect than the corresponding wild-type gene) that are not lethal exhibit a decondensed chromatin structure phenotype, which is most severe on the male X chromosome. The decondensed chromatin structure is concomitant with decreased levels of H3S10 phosphorylation (Deng *et al*., 2008; Wang *et al*., 2001). The role of JIL1 is likely to prevent the formation of a repressive heterochromatin state, complementing the action of the DCC. H3S10 phosphorylation by Aurora B has been demonstrated to displace heterochromatin protein 1 (HP1), which is involved in heterochromatin formation (Hirota *et al*., 2005; Mateescu *et al*., 2004). HP1 recognizes di- and tri-methylated H3K9 via its chromodomain. $H3K9Me_{2/3}$ is an epigenetic mark associated with heterochromatin formation and is catalyzed by the methyltransferase Su(var)3-7 (Jacobs and Khorasanizadeh, 2002; Nielsen *et al*., 2002). The role of HP1 directly antagonizes that of the DCC in dosage compensation of the male X chromosome, promoting transcriptionally repressive chromatin formation and likely functions to regulate the degree of transcriptional upregulation (Kizer *et al*., 2005).

The involvement of both H3S10 phosphorylation and H4K16 acetylation in dosage compensation in *D. melanogaster* is not a novel paradigm in transcriptional activation. Stimulation of mammalian cells with epidermal growth factor (EGF) causes the sequential phosphorylation and acetylation of H3 N-terminal tails and the histone acetyltransferases Gcn5 exhibits a ten-fold increase in substrate specificity for H3S10 phosphorylated H3 N-terminal tails over unmodified N-terminal tails (Cheung *et al*., 2000).

Determinants that govern the specific targeting of the DCC to the male X chromosome are poorly understood. Distribution of the DCC on the male X chromosome, visualized via immunofluorescence studies of the male polytene chromosome from larval salivary glands shows a discontinuous banding pattern of hundreds of sites along the X chromosome. As discussed earlier (section 1.6) it is understood that a subset of approximately thirty to sixty *entry sites* of high affinity for the DCC exist from which the complex spreads in *cis* via the cooperative binding of additional DCCs to sites of weaker affinity (Demakova *et al*., 2003). Chromatin immunoprecipitation studies coupled with high-resolution micro-arrays (ChIP-chip) reveals that genes are preferred targets over intergenic regions and that DCC distribution within the gene is strongest towards the 3' transcribed end with preference for coding sequences (Gilfillan *et al*., 2006). The pattern of distribution is relatively stable between distinct cell lines and development (Legube *et al*., 2006). However, DCC distribution is not a direct marker of dosage compensation as numerous genes that are dosage compensated are not sites of DCC localization (Legube *et al*., 2006). Also the targeting determinant does not appear to be transcriptionally dependent, though localization does appear to be coupled with an active transcriptional state as genes showing the highest levels of bound DCC are the most compensated (Gilfillan *et al*., 2006). A link between DC and transcription seems important as it prevents the ectopic expression of genes that are not being expressed. The observation by Legube *et al*. (2006) that greater than 66% of transcribed X-linked genes are dosage compensated whereas only approximately 15% are bound by MSL1 suggests that the DCC has long-range action. The authors suggest that dosage compensation regions are organized into distinct domains, which have unique transcriptional properties (Legube *et al*., 2006). Alekseyenko *et al*. (2006) carried out a similar study in two male cell lines and in late-stage embryos. They demonstrated an enriched distribution of the DCC complex over expressed genes, and that distribution was enriched towards the 3' end of the gene. Differential binding between the cell lines led them to conclude that sequence alone is not sufficient to explain the pattern of distribution, rather that chromatin factors and/or histone post-translational modifications indicative of active transcription are responsible (Alekseyenko *et al*., 2006).

Association of the hypertranscribed male X chromosome with the nuclear pore complex at the nuclear periphery supports this theory (Mendjan *et al*., 2006). The nuclear pore complex (NPC) is a regulatory complex at the nuclear periphery involved in gene expression and RNA

export (Schmid *et al*., 2006). Interestingly, RNAi knockdown of *Mtor* and *Nup153* expression led to loss of DCC (Mendjan *et al*., 2006). Association of the male X chromosome and physical interaction between the DCC and the NPC supports the theory that dosage compensation regions are organized into distinct domains.

The structure, components and interactions of the DCC are diverse and intricate and whose global action involves epigenetic regulation via the post-translational modification and restructuring of local chromatin to facilitate the two-fold upregulation of the *D. melanogaster* males X chromosome. The components of the DCC complex and synergistic and antagonistic complexes, which began with identification of the male-specific lethal proteins have subsequently been expanded to include Rpd3, JIL1, GAGA factor, supercoiling factor (SCF), ATP-dependent nucleosome remodelers, and components of the nuclear pore. The list is likely to expand as research in this field progresses.

1.6.1 Dosage Compensation Complex in Humans

A H4K16 specific histone acetyltransferase complex containing MOF, MSL1, MSL2, and MSL3 orthologues exist in humans (Marin, 2003; Smith *et al*., 2005). In humans the females dosage compensate by inactivating large regions of one of their X chromosomes, forming a Barr body, to compensate for the males single X chromosome (Lyon, 1961). Unlike its *D. melanogaster* counterpart the human complex is not thought to function in dosage compensation and is not specific to males or the X chromosome. The human complex is responsible for the majority of H4K16 acetylation and is implicated in DNA repair and the G2/M cell cycle arrest checkpoint in response to DNA damage (Gupta *et al*., 2005).

1.7 The Tudor Royal Family and Chromodomains of the Dosage Compensation Complex

Chromodomains were first identified in two known modulators of chromatin structure in *D. melanogaster*, HP1 and Polycomb (PC), which share a “chromo-box” sequence motif. HP1 and PC are involved in constitutive heterochromatin formation and silencing of cell type-specific homeotic gene expression, respectively (Cao *et al*., 2002; Elgin, 1996; Paro *et al*., 1998). Chromodomains (chromatin organization modifier domains) are members of the Tudor Royal family, sharing a conserved three β-stranded core region and are key components of

epigenetic regulation (Maurer-Stroh *et al*., 2003). Chromodomains are a paradigm of how a single structural fold has evolved in diverse proteins to bind distinct targets. Chromodomains are common to nuclear regulators and bind diverse targets including histones, DNA, and RNA (Akhtar *et al*., 2000; Bouazoune *et al*., 2002; Lachner *et al*., 2001). They function as recognition motifs of histone post-translational modifications and facilitate the translation of the histone code into a distinct local chromatin structure via recruiting the appropriate chromatin modulating machinery, as illustrated by Eaf3s chromodomains recognition of the H3K36Me3 epigenetic mark and recruitment of the Rpd3S complex, discussed earlier (section 1.6) (Joshi and Struhl, 2005).

1.7.1 HP1 and Polycomb: Structure and Function

The chromodomain fold of HP1 is a three-stranded antiparallel β–sheet with an α-helix packed against the β-strands (Jacobs and Khorasanizadeh, 2002; Nielsen *et al*., 2002). The fold is structurally homologous to the DNA binding protein Sso7d, a chromosomal binding protein from the hyperthermophilic archaebacteria *Sulfolobus solfataricus* (Gao *et al*., 1998). HP1 binds methylated H3K9, an epigenetic marker associated with gene silencing. H3K9 methylation is catalyzed by the histone methyltransferase Su(var)3-9 (Lachner *et al*., 2001). The crystal structure of the HP1 chromodomain bound to a H3K9$Me_{2/3}$ peptide reveals that the modified histone tail binds to HP1 via insertion of the tail as a β-strand completing a β-sandwich fold (Figure 1.8). The methyl-ammonium group is buried in a hydrophobic cage, formed by the side chains of Y24, Y48, and W45 (Figure 1.9).

Polycomb specifically binds tri-methylated H3K27. Methylation of H3K27 is catalyzed in *D. melanogaster* by an Enhancer of Zeste (E[z]) and Extra Sex combs (ESC) complex (Czermin *et al*., 2002). The structures of the PC and HP1 chromodomains are highly similar. The H3K27Me_3 binds in a similar mechanism, the peptide incorporating as a β-strand and the methyl-ammonium group buried in a hydrophobic cage formed by the side chains of Y26, W47, and W50. When the structures are overlaid the chromodomain and peptide Cα positions only deviate by a pair-wise root-mean-squared-difference (rmsd) of 1.1 Å and 0.45 Å, respectively (Min *et al*., 2003). Specificity of dPC for H3K27Me_3 appears to be achieved via dimerization of the dPC chromodomain. HP1 does not dimerize. Dimerization of dPC positions the H3K27Me_3 N-terminal tails in close proximity with each other across the dimer interface allowing for

histone-histone interaction between adjacent histone H3 tails. If H3K9Me were to bind in a similar fashion Arg 2 and Lys 4 would form significant steric conflicts (Cowell and Austin, 1997; Min *et al*., 2003).

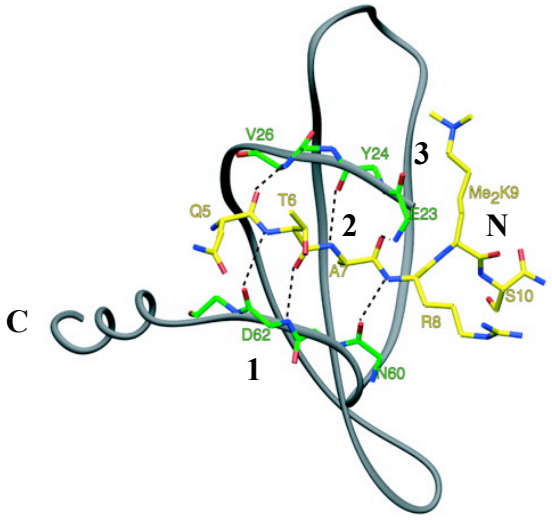


Figure 1.8: Structure of the HP1 chromodomain. The chromodomain fold of HP1 is a three-stranded antiparallel β–sheet with an α-helix perpendicular to the β-strands. A backbone trace reveals incorporation of the H3K9Me$_2$ peptide as a β-strand completing the β-sandwich. Figure taken from (Jacobs and Khorasanizadeh, 2002).

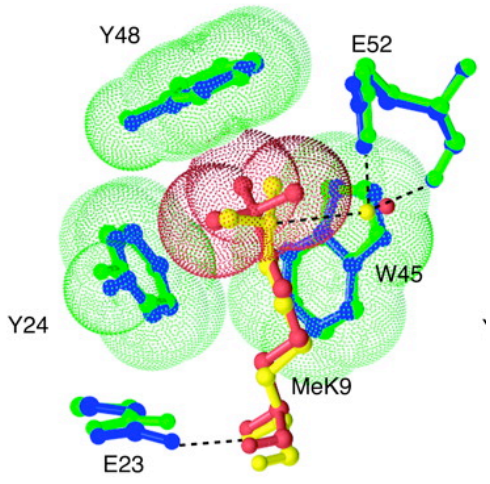


Figure 1.9: Hydrophobic binding pocket of HP1. The H3K9Me$_3$ (red peptide, green structure) structure is shown overlaid with that of the H3K9Me$_2$ (yellow peptide, blue structure) structure. Van der Waals radii are shown for the H3K9Me$_3$ structure. Figure taken from (Jacobs and Khorasanizadeh, 2002).

1.7.2 Cooperativity within the Tudor Royal Family

The concerted action of two domains in the recognition of a specific methyl-lysine mark is also observed in the demethylase JMJD2A, a Jmjc-domain family member, whose C-terminal double Tudor domains recognizes and binds histones methylated at H3K4 or H3K20 with similar affinities (K_d=0.50μM and K_d=0.40μM, respectively) (Huang *et al*., 2006; Lee *et al*., 2008; Tsukada *et al*., 2006). The structure of the double tudor domains complexed with a H3K4Me_3 peptide reveals two canonical tudor domains interdigitated by two β-strands, β2 and β3, which both domains share (Figure 1.10). The recent publication of the JMJD2A-H3K20Me_3 complex structure reveals that binding of the two peptides is distinct. Despite methyl-lysine binding involving the same aromatic cage residues the peptide orientation is opposite and contacts different surfaces (Lee *et al*., 2008).

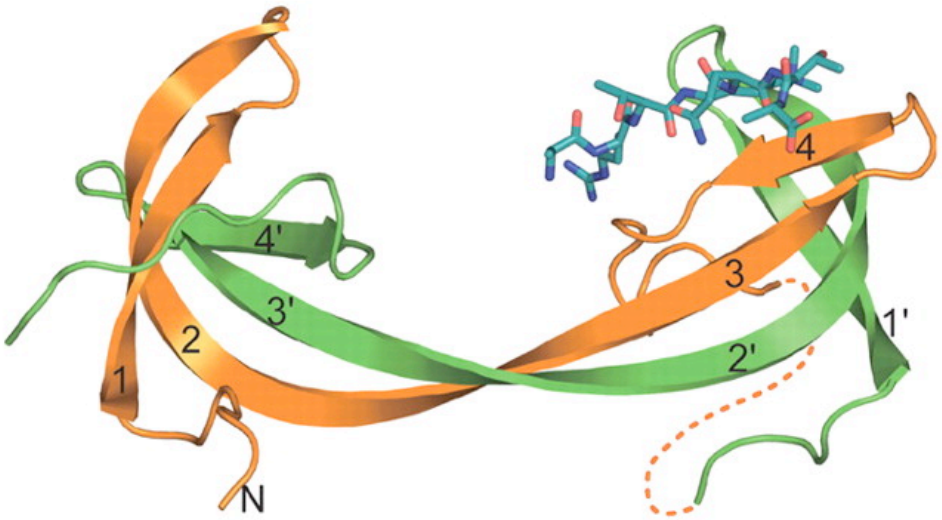


Figure 1.10: Structure of the JMJD2A double Tudor domains in complex with H3K4Me_3 (blue). The double Tudor domains of are interdigitated by β-strands β2 and β3. Figure taken from (Huang *et al*., 2006).

The H3K4Me$_3$ peptide is predominantly bound by the second domain. Like the chromodomains of HP1 and PC the methyl-ammonium group is buried in a hydrophobic cage. The hydrophobic cage is formed by W967 and Y973, of the second domain and F932 of the first domain. Specificity is conferred via side-chain interactions between the second domain and the peptide (Figure 1.11). Unlike HP1 and PC the histone tail is not incorporated as a β-strand (Huang *et al*., 2006).

53BP1 (p53 binding protein 1), involved in the DNA damage check-point signal also contains a double Tudor domain but they are not interdigitated and form two separate domains. The NMR structure reveals the tudor domains to have a β-barrel structure with close evolutionary relationship to the tudor domain of Survival of Motor Neuron (SMN), which is thought to bind a RG rich motif (Charier *et al*., 2004; Selenko *et al*., 2001; Sprangers *et al*., 2003). The closely related malignant brain tumor repeats of human L3MBTL1 have been shown to bind a monomethylated lysine histone substrate, suggesting that the SMN domain may have a similar function (Li *et al*., 2007b).

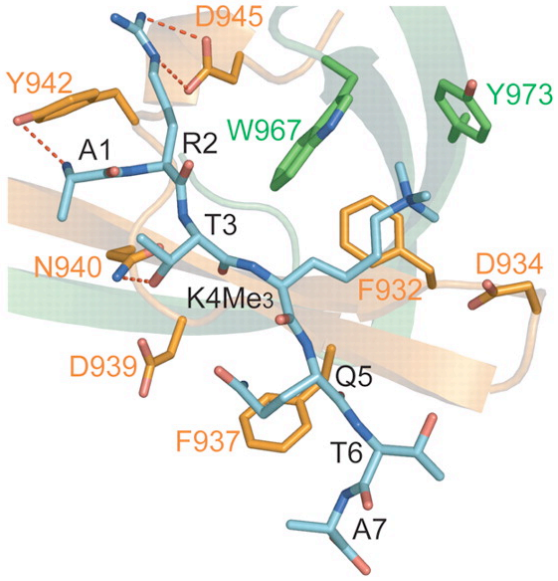


Figure 1.11: Binding of H3K4Me$_3$ by the double Tudor domains of JMJD2A. The methyl-ammonium group is bound by W967 and Y973, of the second domain and F932 of the first domain. Specificity is conferred via side-chain interactions between the second domain and the peptide. Figure taken from (Huang *et al*., 2006).

Interestingly the double Tudor domain of 53BP1 binds H4K20Me$_2$. The interaction involves Van der Waals π-cation interactions between the dimethyl-lysine and a hydrophobic cavity formed by the side chains of W1495, Y1502, F1519, and Y1523. Specificity for dimethyl-lysine over trimethyl-lysine is conferred via a salt-bridge between the carboxylate group of D1521 and the methyl-ammonium group of H4K20Me$_2$. A hydrogen bond between the carboxylate group of D1521 and the amino proton of H4K20Me$_2$ is also likely to confer specificity. Trimethyl-lysine does not have the amino hydrogen and would therefore not be able to form the hydrogen bond (Botuyan *et al*., 2006).

The double chromodomains of the chromatin remodeling protein CHD1 (chromo-ATPase/helicase-DNA binding domain 1) of the SAGA and SLIK acetyltransferase complexes also act cooperatively to bind their target, H3K4Me (Flanagan *et al*., 2005; Pray-Grant *et al*., 2005). The crystal structure of the CHD1 double chromodomain complexed with an H3K4Me$_3$ peptide reveals the two chromodomains are linked by a helix-loop-helix structure and that both chromodomains have similar secondary structure to that of HP1 and PC. The H3K4Me$_3$ group is bound at the interface of the two chromodomains. Two aromatic residues, W64 and W67 from chromodomain 1 interact with the methyl-ammonium group unlike HP1 and PC, which use the canonical three aromatic residue cage. The π-cation interaction between the methyl-ammonium group and W64 and W67 is further stabilized by a π-cation interaction between R2 of the H3K4Me$_3$ peptide and W67. This interaction is proposed to confer specificity for methylated H3K4 (Figure 1.12). Chromodomain 2 of CHD1 has only one of the two aromatic residues and is therefore unable to bind the methyl-ammonium group. In both HP1 and PC, the H3 peptide inserts as a β-strand completing the chromodomains β-sandwich fold. An insertion novel to chromodomain 1 of CHD1 plays this role, and thus prevents peptide binding in a manner homologous to HP1 or PC. Rather the H3K4 peptide is bound at the domain interface (Flanagan *et al*., 2005). It is interesting to note that binding of HP1 and the CHD1 chromodomains to their respective targets is abrogated by phosphorylation of H3S10 or H3T3, respectively (Flanagan *et al*., 2005; Hirota *et al*., 2005). The double chromodomain of the CHD1 isoform Mi-2 in *D. melanogaster* does not bind H3K4Me$_3$, rather it has been shown to bind DNA (Bouazoune *et al*., 2002), though some believe this to be unlikely.

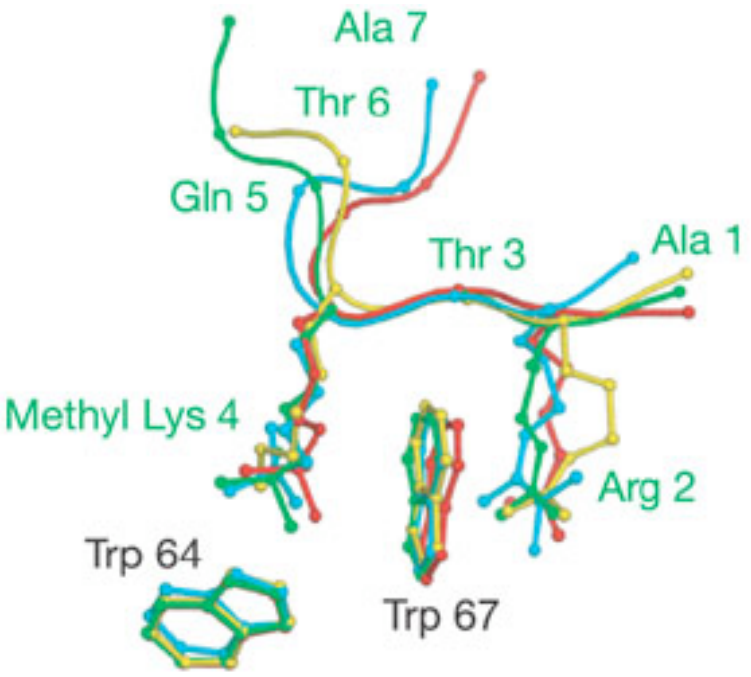


Figure 1.12: Binding of H3K4Me by the double chromodomain of CHD1. The H3 peptide is bound by two aromatic residues, W64 and W67, which are further stabilized by interaction of R2 with W67. H3K4Me (yellow), $H3K4Me_3$ (green), $H3K4Me_3R2Me_2$ (cyan), $H3K4Me_3T3ph$ (red). Figure taken from (Flanagan *et al*., 2005).

1.7.3 The Chromodomains of MOF and MSL3

MOF and MSL3, two components of the DCC contain chromodomains (Figure 1.7). To date no direct evidence exists suggesting that MOF binds a methyl-lysine epigenetic mark. The MSL3 chromodomain has been demonstrated to recognize $H3K36Me_3$ (Bell *et al*., 2008; Larschan *et al*., 2007). The chromodomain of Eaf3, an MSL3 paralogue in *S. cerevisiae* has also been demonstrated to recognize methylated H3K36 (Joshi and Struhl, 2005). Sequence alignment of the MOF and MSL3 chromodomains with HP1 and PC reveals that MSL3 exhibits the three conserved aromatic residues associated with forming the hydrophobic cage that binds the methyl-ammonium group of a methylated lysine residue in HP1 and PC. MOF however does not exhibit the three conserved aromatic residues (Figure 1.13).

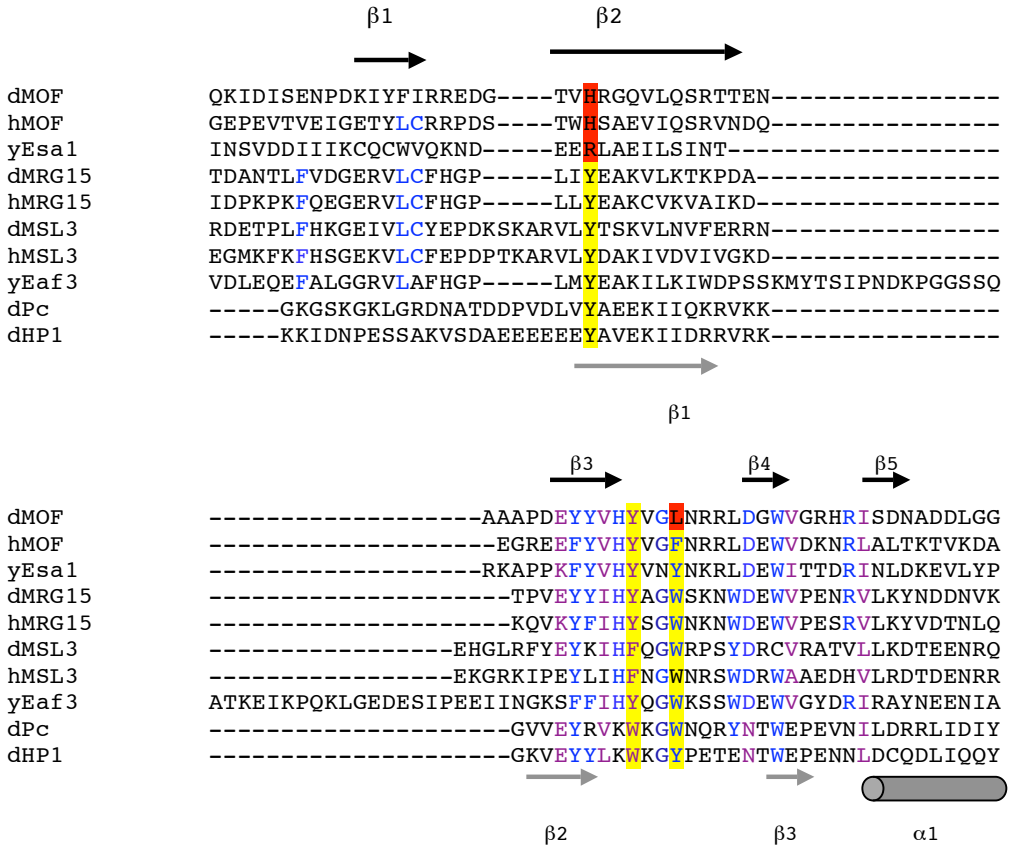


Figure 1.13: Sequence alignment of the MOF and MSL3 chromodomains of the DCC with their homologues yEsa1, hMOF, hMSL3, yEaf3, dMRG15, hMRG15 and archetypal chromodomains HP1 and PC. Secondary structure elements are shown in grey for HP1 and black for MOF. Residues involved in forming the methyl ammonium binding hydrophobic cage are highlighted in yellow. Conserved residues are shown in blue, broadly conserved residues in purple. Important residues that differ between MOF and MSL3 are shown in red.

Akhtar *et al*. (2000) provided preliminary evidence supporting the involvement of the MOF and MSL3 chromodomains in binding RNA, and that their association with the DCC is RNAse sensitive (Akhtar *et al*., 2000; Smith *et al*., 2000). Some doubt remains whether these chromodomains actually bind RNA. The notion that the MSL3 chromodomain (2-91) is sufficient for binding an oligonucleotide directly conflicts with previous findings that nucleotide binding is localized to residues 140-259 (Morales *et al*., 2005). The possibility of binding an RNA molecule however does not rule out that one or both may also bind a modified

histone tail. It is not known if the MOF or MSL3 chromodomains bind their putative targets as a monomer or as a dimer.

During the course of this research the NMR structure of the MOF chromodomain was published by Nielsen *et al*. (2005). The structure, comprising residues 383-433 consists of five β-strands forming a β-barrel leading the authors to coin the name chromo-barrel domain. The barrel consists of two β-sheets of three strands with β2 spanning both sheets. β5, β2 and β1 on one side and β2, β3 and β4 on the other. The structure of the chromo-barrel is significantly different to the archetypal chromodomains of HP1 and PC. The chromo-barrel has no C-terminal α-helix, has two extra β-strands and the chromo-box motif is located on the β3-β4 loop, as opposed to the β2-β3 loop. The most significant difference is revealed when the MOF chromo-barrel and HP1 chromodomain complexed to an H3K4Me peptide are superimposed. β1 of the MOF chromo-barrel occupies the peptide-binding site seen in HP1 and PC, therefore MOF cannot bind a histone tail in a homologous manner (Nielsen *et al*., 2005) (Figure 1.14). This however does not rule out participation in peptide binding as the same situation is observed in chromodomain 1 of the CHD1 double chromodomain, discussed earlier (section 1.7.2) (Flanagan *et al*., 2005).

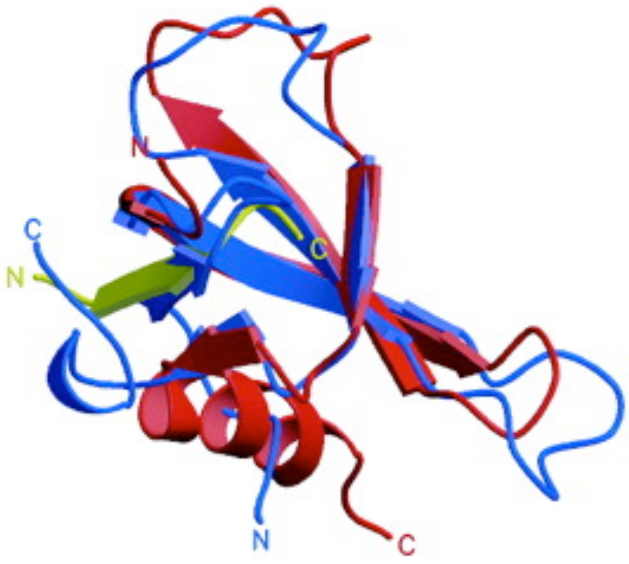


Figure 1.14: Superposition of the MOF chromo-barrel domain (blue) with the archetypal chromodomain of HP1 (red) complexed with a H3K4Me peptide (green). Figure taken with permission from (Nielsen *et al*., 2005).

The crystal structure of the N-terminal chromodomain of the *H. sapiens* MRG15 (MORF4-related gene on chromosome 15) transcription factor to 2.2 Å resolution was published soon after the publication of the solution NMR structure of the *D. melanogaster* MOF (371-440) chromo barrel domain (Zhang *et al*., 2006). The MRG15 transcription factor is found in the MAF1 (MRG-associated factor 1), MAF2, and mammalian NuA4/Tip60 complex, which function in chromatin remodeling and transcriptional regulation (Garcia *et al*., 2007; Pardo *et al*., 2002). The MRG15 chromodomain is closely related to the *D. melanogaster* MSL3 (2-93) and the *S. cerevisiae* Eaf3 chromodomain. Like its *D. melanogaster* and *S. cerevisiae* homologues, the MRG15 protein is likely to bind a H3K36Me peptide via its chromodomain (Zhang *et al*., 2006). Like Eaf3 it is also likely to associate independently with the HDAC Sin3 and NuA4/Tip 60 HAT complexes (Cai *et al*., 2003; Doyon *et al*., 2004; Yochum and Ayer, 2002). The MRG15 chromodomains structure adopts a β-barrel closely resembling that of the MOF chromo-barrel domain, with a long C-terminal α-helix (Figure 1.15).

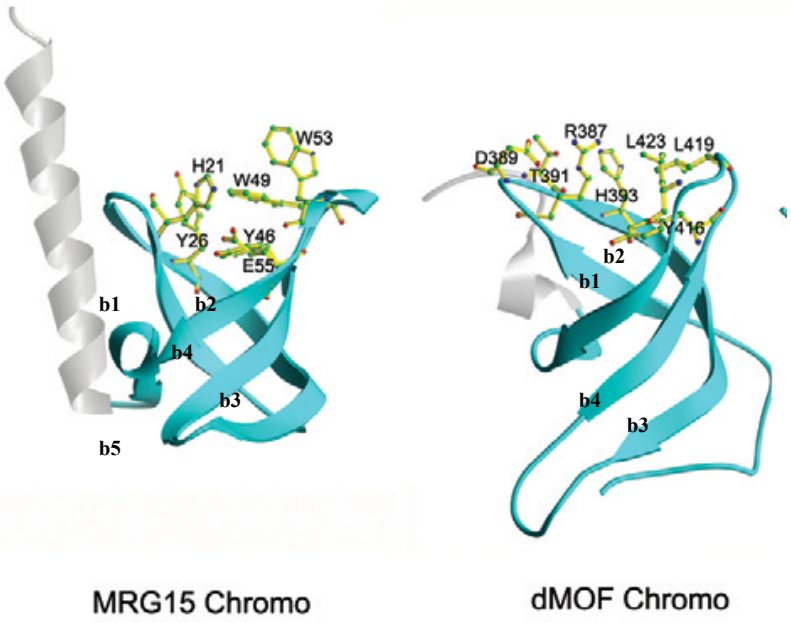


Figure 1.15: Crystal structure of the *H. sapiens* MRG15 chromo barrel domain (left) aligned next to the *D. melanogaster* MOF chromo-barrel domain (right). Figure taken from (Zhang *et al*., 2006).

Interestingly, like the MOF chromo-barrel domain, β1 of the MRG15 chromodomain occupies the peptide-binding site seen in HP1 and PC therefore the MRG15 chromodomain is unlikely to bind H3K36Me in a homologous manner (Figure 1.16). Binding of the methyl-lysine is likely to involve a hydrophobic pocket formed by three aromatic residues, Tyr26, Tyr46, and Trp49 (Zhang *et al*., 2006).

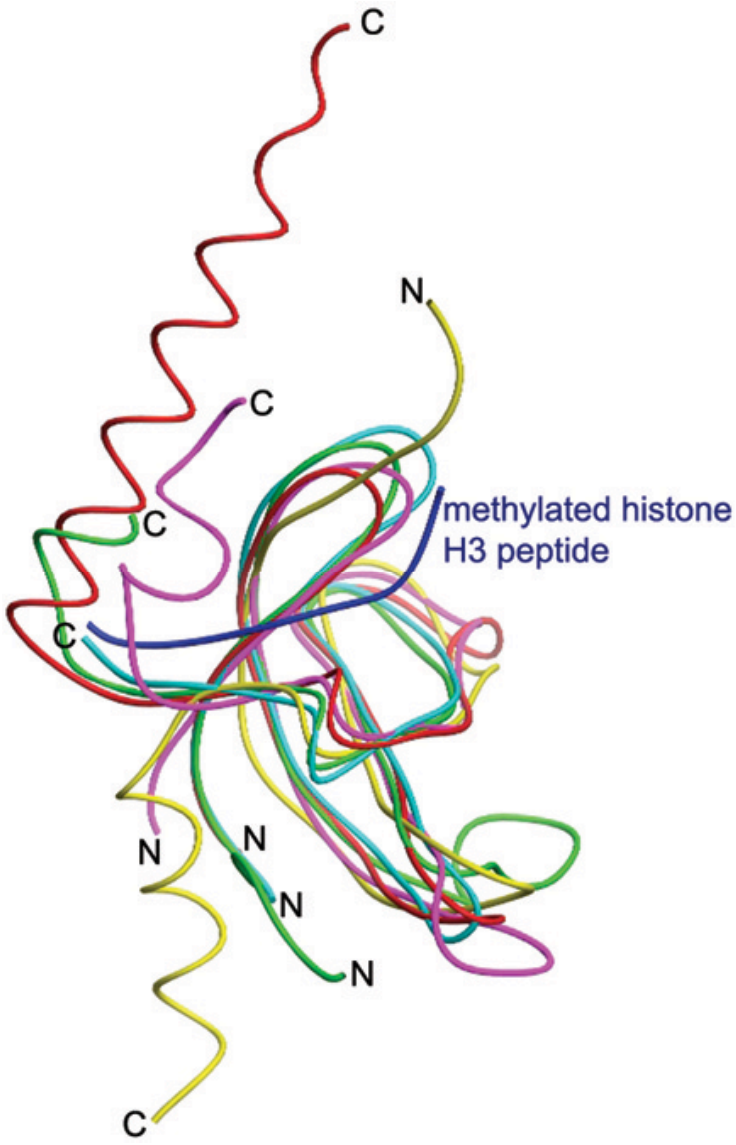


Figure 1.16: Structural overlay of the *H. sapiens* MRG15 chromo-barrel domain (red) with the archetypal HP1 chromo domain (yellow), the SMN Tudor domain (cyan), the Dnmt3b PWWP domain (green), and the *D. melanogaster* MOF chromo-barrel domain (magenta) demonstrating structural conservation. Figure taken from (Zhang *et al*., 2006).

It has been suggested that the *D. melanogaster* MOF (371-440) and MSL3 (2-91) chromodomains bind RNA (Akhtar *et al*., 2000). Akhtar *et al*. (2000) demonstrated a region of MOF (352-518) containing the chromo-barrel domain was responsible for RNA binding via electromobility shift assays using a roX1 RNA fragment. Interestingly, in a more recent study no RNA binding activity was attributable to the chromodomain of the closely related MYST HAT Esa1 (17-89), which also adopts a β-barrel fold most closely related to the SMN tudor domain (Shimojo *et al*., 2008). However, an N-terminally extended Esa1 construct (1-89) bound a poly(U) oligonucleotide. The solution NMR structure reveals that the N-terminal addition forms a β-sheet that acts as a knot for the chromo-barrel domain. Mutations within this N-terminal extension lost RNA binding ability and were lethal (Shimojo *et al*., 2008). A similar investigation contradicted the assertion that the *D. melanogaster* MSL3 chromodomain binds RNA (Morales *et al*., 2005). Morales *et al*. (2005) failed to identify any RNA binding attributable to the chromodomain itself, rather they demonstrated that RNA binding was localized to a basic region (140-259) between the chromodomain and the MRG domain. Interestingly, deletion of the MSL3 chromodomain had no visible effect upon targeting to the X chromosome on global acetylation of H4K16. The MSL3 chromodomain is, however important for dosage compensation and male viability (Buscaino *et al*., 2006).

1.8 Thesis Objectives

Chromodomains are members of the Tudor Royal family, sharing a conserved three β-stranded core region and are a paradigm of how a single structural fold can evolve in diverse proteins to bind distinct targets. Chromodomains typically function as recognition motifs of methylated lysine residues in histone N-terminal tails and facilitate the translation of the histone code into a distinct local chromatin structure via recruiting the appropriate chromatin modulating machinery.

Two subunits, the MYST HAT MOF and MSL3 of the DCC of *D. melanogaster*, a ribonucleoprotein complex associate with the two-fold transcriptional upregulation of the male X chromosome exhibit chromodomains. At the commencement of this research there was no published structural information on the MOF and MSL3 chromodomains or any direct evidence of what their function may be other than tentative evidence that they may bind RNA.

The goal of this research was therefore to investigate the structure of the two respective chromodomains by NMR and/or X-ray crystallography via the expression and purification of recombinant protein expressed in a bacterial expression system. Structural solution of the MOF and MSL3 chromodomains will advance our understanding of chromodomains in epigenetic processes, specifically in the DCC.

2.0 Materials and Methods

2.1 Cell Lines

The *E. coli* strain XL1 (ampR) from Stratagene was used for transformation and plasmid preparation. The *E. coli* strains Rosetta 2 (DE3) pLysS and BL21 (DE3) from Invitrogen and Novagen, respectively were used to express the recombinant protein of interest.

2.2 Materials

Chemicals and reagents, media, equipment, enzymes and their buffers, and DNA mentioned in the common methods section are listed below.

Table 2.1: List of chemicals and reagents, media, enzymes and enzyme buffers, equipment, chromatography resin, and DNA used in the experimental procedures.

Item	Supplier
Chemicals and Reagents	
Acrylamide:bis [37.5:1]	OmniPur
Agar	BD Beckton Dickinson Co.
Agarose	OmniPur
Ammonium sulfate	Sigma-Aldrich
Ampicillin	Sigma-Aldrich
β-mercaptoethanol	Sigma-Aldrich
Bromophenol blue	Sigma-Aldrich
Bovine serum albumin	Sigma-Aldrich
^{13}C-D-glucose	CIL Cambridge Isotope Laboratories
2-(Cyclohexylamino)ethanesulfonic acid	Sigma-Aldrich
Chloramphenicol	Sigma-Aldrich

Coomassie brilliant-blue R250	Sigma-Aldrich
Dithiothreitol	Sigma-Aldrich
Ethylenediaminetetraacetic acid	Sigma-Aldrich
Ethanol	Sigma-Aldrich
Ethidium bromide	Sigma-Aldrich
Glacial acetic acid	EMD
Glutathione	Sigma-Aldrich
Glycerol	Sigma-Aldrich
Glycine	Sigma-Aldrich
L-Isoleucine	General Biochemicals
L-Leucine	Sigma-Aldrich
L-Lysine	General Biochemicals
L-Phenylalanine	ICN Biochemicals
L-Seleno-Methionine	Sigma-Aldrich
L-Threonine	Nutritional Biochemicals Corp.
L-Valine	General Biochemicals
Magnesium acetate	Sigma-Aldrich
Magnesium sulfate	Sigma-Aldrich
Methanol	Sigma-Aldrich
^{15}N-Ammonium chloride	CIL Cambridge Isotope Laboratories
Phenylmethylsulphonyl fluoride	Sigma-Aldrich
Potassium acetate	Sigma-Aldrich
Potassium chloride	Sigma-Aldrich
Potassium iodide	Sigma-Aldrich
Potassium phosphate (mono-basic)	Sigma-Aldrich
Sodium Dodecyl Sulfate	Sigma-Aldrich
Sodium acetate	Sigma-Aldrich
Sodium bromide	Sigma-Aldrich
Sodium chloride	Sigma-Aldrich
Sodium phosphate (di-basic)	Sigma-Aldrich
Tris(2-carboxyethyl)phosphine hydrochloride	Sigma-Aldrich

Tris-acetate	Sigma-Aldrich
Triton X-100	Sigma-Aldrich
Trizma base	Sigma-Aldrich
Tryptone	BD Beckton Dickinson Co.
Yeast Extract	BD Beckton Dickinson Co.

Media

Agar	BD Beckton Dickinson Co.
Tryptone	BD Beckton Dickinson Co.
Yeast Extract	BD Beckton Dickinson Co.

Enzymes and Enzyme Buffers

10X Buffer Tango	Fermentas
10X DNA Ligase buffer	Fermentas
*Bam*HI	Fermentas
*Bgl*II	Fermentas
*Eco*RI	Fermentas
Pfu DNA polymerase	Fermentas
*Sal*I	Fermentas
Taq DNA polymerase	Fermentas
Trypsin	Sigma-Aldrich

Equipment

Äkta FPLC	Amersham Biosciences
Äkta Prime	Amersham Biosciences

Pipettes	Gilson
Agarose gel apparatus	Owl Separation Systems
Mighty Small II SDS PAGE apparatus	Amersham Biosciences
Power supply (model 300)	VWR

Resin

SourceQ Anion exchange resin	Amersham Biosciences
SourceS Cation exchange resin	Amersham Biosciences
Glutathione Sepharose 4B	Amersham Biosciences

DNA

D. melanogaster MSL3 and MOF cDNA	Prof. Scott, M. J., Massey University, Palmerston North, New Zealand.
H. sapiens MSL3 and MOF cDNA	Dr. Edwin Smith, Stowers Institute for Medical Research, Kansas City
Synthetic oligonucleotides	Sigma-Aldrich

2.3 Common Methods

2.3.1 Agarose Gel Electrophoresis

DNA samples were mixed with a final concentration of 1X DNA loading buffer (0.04% [w/v] bromophenol blue, 5% [v/v] glycerol) from a 10X stock solution. 10-20 μL was then loaded onto a 2-5% agarose gel. The agarose gel consisted of 2-5% w/v agarose in 1X TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.0) from a 50X stock, and 0.5 μg/mL ethidium bromide.

The DNA samples were resolved on the agarose gel at 100 V for 10-20 min in 1X TAE. Bands were visualized using a UV lamp at 365 nm.

2.3.2 SDS Polyacrylamide Gel Electrophoresis

Protein samples were mixed with a final concentration of 1X Laemmli loading dye (80 mM Tris-HCl, pH 6.8, 2% [w/v] SDS, 10% [v/v] glycerol, 0.2 μg/mL bromophenol blue, 5% [v/v] β-mercaptoethanol). 10-20 μL was then loaded onto a 4% stacking SDS polyacrylamide gel (150 mM Tris-HCl, pH 6.8, 0.1% [w/v] SDS, 4% [v/v] 40% acrylamide:bis [37.5:1], 0.5% [v/v] APS, 0.1% [v/v] TEMED) with a 17% resolving SDS polyacrylamide gel (450 mM Tris-HCl, pH 8.8, 0.1% [w/v] SDS, 17% [v/v] 40% acrylamide:bis [37.5:1], 0.5% [v/v] APS, 0.05% [v/v] TEMED). All gels consisted of a 4% stacking gel and a 17% resolving gel. The protein samples were resolved on the SDS polyacrylamide gel at 20 mA in Tris-Glycine Laemmli running buffer (25 mM Tris-HCl, 200 mM glycine, 0.1% [w/v] SDS) until the dye band had run off. Bands were visualized by soaking the gel in a Coomassie staining solution (40% [v/v] methanol, 10% [v/v] glacial acetic acid, 0.1% [w/v] Coomassie brilliant-blue R250) followed by soaking in a destain solution (40% [v/v] methanol, 10% [v/v] glacial acetic acid).

2.3.3 PCR amplification

In order to clone the gene fragment of interest, it was first amplified by PCR. The PCR reaction solution contained 1X reaction buffer (200 mM Tris-HCl, pH 8.8, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 1% [v/v] Triton X-100, 1 mg/mL BSA, 20 mM $MgSO_4$), 200 μM dNTPs, 200 nM 5' primer, 200 nM 3' primer, 1 μL *Pyrococcus furiosus* (Pfu) DNA polymerase and approximately 1 ng of template DNA. The template was amplified via thirty cycles using a BioRad MyCycler thermal cycler (Figure 2.1). Primers are shown in Table 2.2.

2.3.4 Ethanol Precipitation of PCR DNA

To 100 μL of PCR product, 10 μL of sodium acetate (3 M, pH 5) was added followed by 300 μL of 95% ethanol. After storage at -20°C for 10 min, the sample was then centrifuged at 17418 x g (r_{max},12000 rpm, Beckman JA-20 rotor) for 10 min at 4°C to pellet the DNA. The supernatant was discarded and the pellet washed with 200 μL of 70% ethanol. The pellet was

dried by rotary-evaporation and redissolved in 20 μL of TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA).

2.3.5 Restriction Endonuclease Digestion

The DNA sample to be digested was first quantified approximately by running a known volume on an agarose gel (described in section 2.3.1) along with DNA ladder markers of known size and concentration and comparing their relative intensities by ethidium bromide fluorescence. A reaction mixture containing 2X buffer Tango (3.3 mM Tris-acetate, pH 7.9, 1 mM magnesium acetate, 6.6 mM potassium acetate, 0.01 mg/mL BSA), sample DNA and 0.1 μL (one unit) of the appropriate restriction endonuclease (10 units/μL) per μg of sample DNA was incubated for one hour at 37°C.

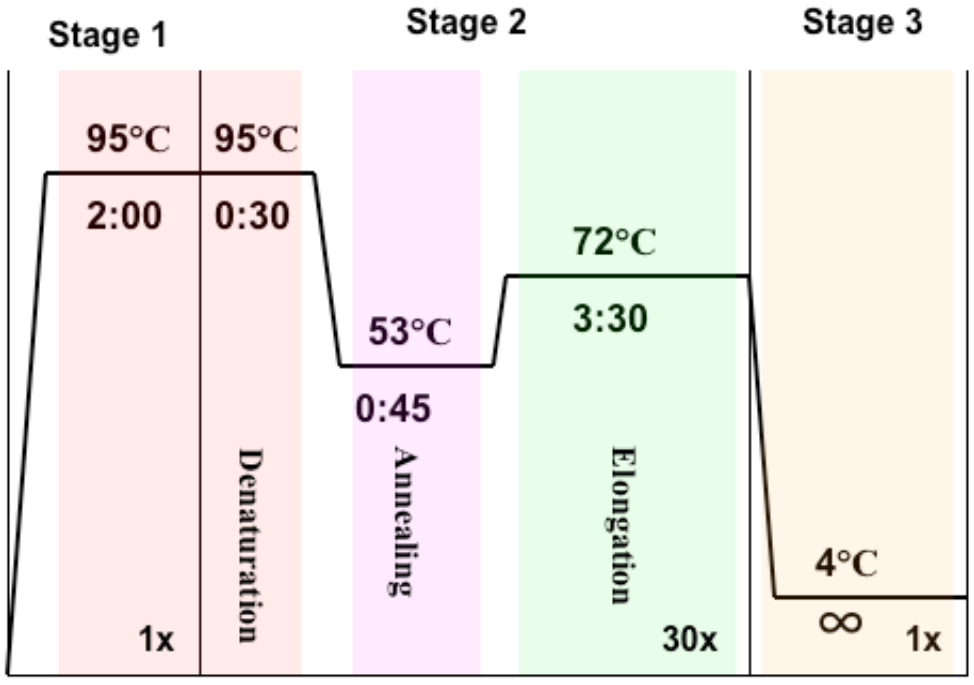


Figure 2.1: Polymerase chain reaction protocol used for amplification of the specific segment of DNA required. Time is shown in minutes. Stage 1 is followed by thirty cycles of stage 2, and then held indefinitely at 4° during stage 3.

Table 2.2: Primers for cloning the *H. sapiens* MOF and MSL3, *S. pombe* SWI6, and *D. melanogaster* MOF and MSL3 chromodomains. The restriction endonuclease site is shown in red and the stop codon in blue. The forward primer is listed first, the reverse primer second. The *D. melanogaster* MOF and MSL3 chromodomains were cloned by Joanne Guenther (University of Saskatchewan) and Bradley McClellan (Massey University), respectively. Ewa Kerc (University of Saskatchewan) contributed to the cloning of *S. pombe* SWI6 and *H. sapiens* MOF and MSL3 chromodomains.

Species		Sequence	Restriction Endonuclease
D. melanogaster			
MOF (371-440)	5'	GAC CTG GGA TCC CAA AAG ATC GAT ATA AGC GAA AAT CCC	*Bam*HI
	5'	GAC GAG GAA TTC CTA GCC CAA GTC ATC CGC ATT GTC	*Eco*RI
MSL3 (2-96)	5'	GAC CTG GGA TCC ACG GAG CTA AGG GAC GAG ACA	*Bam*HI
	5'	GAC GAG GAA TTC CTA ACC ACG AAT TTG CAG CTT GGC	*Eco*RI
MSL3 (2-91)	5'	GAC CTG GGA TCC ACG GAG CTA AGG GAC GAG ACA	*Bam*HI
	5'	GAC GAG GAA TTC CTA CTT GGC AGC TTC GGC CAG	*Eco*RI
H. sapiens			
MOF (21-93)	5'	GCC GAG GGA TCC GGC GAG CCG GAA GTC ACG	*Bam*HI
	5'	GAG GCC GTC GAC CTA CTT CTG TAC AGC ATC CTT CAC TGT	*Sal*I
MSL3 (2-93)	5'	GAC CTG GGA TCC AGC GCG AGC GAG GGC ATG AAA TTT	*Bam*HI
	5'	GAC GAG GAA TTC CTA GCG AGC TAC AGC TTT TCT TGC CAA	*Eco*RI
S. pombe			
SWI6 (77-135)	5'	GAC CTG GGA TCC GAA GAG GAT GAA TAT GTT GTA GAA AAG	*Bam*HI
	5’	GAC GAG GAA TTC TCA CGC CTC GTT CGC CAG	*Eco*RI

2.3.6 Agarose Gel Purification of DNA

DNA was purified by loading onto a 1-5% agarose gel (lower percentage for larger fragments) and resolved as previously described (section 2.3.1). The band corresponding to the PCR amplified fragment was then excised from the gel in the smallest amount of agarose possible and weighed. The DNA was extracted from the agarose using the Qiagen Gel Extraction Kit, following the instructions contained. The DNA was eluted with 30 μL of TE buffer.

2.3.7 DNA Ligation

The cloned fragment of interest was PCR amplified and along with the desired vector individually digested with the appropriate restriction endonucleases and gel-purified as described earlier (section 2.3.5-6). The ligation of the insert and vector was carried out in the smallest volume possible, typically 10 μL containing 2 μL vector, 1 μL insert (at a 4:1 insert:vector molar ratio), 1 μL T4 DNA ligase, 2 μL 5X reaction buffer, and 4 μL double-distilled H_20 (ddH_2O). The reaction was allowed to incubate at room temperature (approximately 26^oC) overnight (O/N).

2.3.8 Competent Cells

Competent cells were prepared by inoculating 5 mL of LB with a single XL1 colony and incubated at 37°C O/N. 100 mL of sterile LB was then inoculated with 1 mL of the O/N, and grown until an $OD_{(600nm)}$ of 0.6 was reached. Cells were then harvested by centrifugation at 7741 x g (r_{max}, 8000 rpm, Beckman JA-20 rotor) for 10 mins. Keeping the cells on ice at all times, the cells were resuspended in 20 mL of ice-cold 100 mM $MgCl_2$ solution and incubated on ice for 20 min. The cells were then harvested by centrifugation at 7741 x g and resuspended in 2 mL of a 100 mM $CaCl_2$, 14% glycerol solution. The cells were immediately snap-frozen in 100 μL aliquots and stored at -80°C. The transformation efficiency of the competent cells, expressed as the number of colonies obtained from 1 μg of DNA, was approximately 1 x 10^6 cells/μg DNA.

2.3.9 Transformation of *Escherichia coli* cell lines

The 10 μL DNA ligation reaction or 1-2 μL of dilute recombinant plasmid was added to 50 μL of competent *E. coli* XL1 cells. The cells were then cold-shocked by placing on ice for 45-60 min and then plated onto pre-warmed (37°C) LB agar plates with ampicillin (100 μg/mL). The plates were then incubated at 37°C O/N.

For the transformation of expressing strains, 1-2 μL (approximately 20 ng) of the appropriate plasmid was added to 50 μL of competent *E. coli* Rosetta 2 cells. The cells were then cold-shocked by placing on ice for 45-60 min and then plated onto pre-warmed (37°C) LB agar plates with ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL).

2.3.10 Recombinant Screening

Colony PCR was employed to screen recombinant colonies using PCR primers used to amplify the respective inserts. A PCR reaction was set up identically to the cloning PCR reaction but using *Thermus aquaticus* (Taq) polymerase and its respective buffer. Instead of template DNA the reaction mixture was inoculated with a colony from the transformation agar plate. Care was taken not to remove the entire colony. A separate reaction was set up for each colony. A standard PCR cycle was run and 5 μL of each reaction, mixed with 5 μL of 2X DNA loading buffer, was loaded on a 1% agarose gel. The cloned PCR insert was loaded as a control. A band exactly the same size as the control was observed for positive colonies. Positive colonies were re-streaked onto separate ampicillin agar plates.

2.3.11 Plasmid Preparation

For positive colonies, 5 mL of LB broth containing 100 μg/mL ampicillin was inoculated with a separate colony and grown O/N at 37°C with shaking. The cells were harvested by centrifugation at 15008 x g (r_{max}, 10,000 rpm, Beckman JLA-16.250 rotor) for 1 min. The supernatant was discarded. The cells were lysed and the plasmid purified using the Qiagen Plasmid Mini-Prep kit, following the instructions contained within. To verify the correct insert, approximately 1 μg of the respective plasmid was single and double restriction digested and run on an agarose gel with digested vector and insert as controls. The single cut plasmid should yield one band running slower than cut vector, while the double cut plasmid should yield two bands equal in size to the cut vector and the insert, respectively. The inserts were also sequenced at the Plant Biotechnology Institute DNA sequencing facility at the University of Saskatchewan for further confirmation.

2.3.12 Expression of Recombinant Protein

Recombinant protein was expressed in Rosetta 2 *E. coli* cells (Invitrogen). A 5 mL aliquot of LB containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol was inoculated with a single colony and grown O/N at 37°C. A 6 L flask containing 2 L of LB containing the appropriate antibiotics was then inoculated with 1 mL of the O/N and grown to the appropriate optical density (OD_{600nm}) at 37°C. The cells were then induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and either grown at 37°C or 25°C O/N, depending upon the

Table 2.3: Expression conditions for the *S. pombe* SWI6, *H. sapiens* and *D. melanogaster* MOF and MSL3 chromodomains. Induction with 0.1 mM IPTG followed by incubation at 25°C for 16 hr was optimal for all constructs.

Species	OD
D. melanogaster	
MOF (371-440)	0.2
MSL3 (2-96)	0.6
MSL3 (2-91)	0.6
H. sapiens	
MOF (20-93)	0.6
MSL3 (2-93)	0.8
S. pombe	
SWI6 (77-135)	0.8

optimal expression conditions determined for that recombinant protein. Optimal conditions are summarized in Table 2.3.

2.3.13 Cell Lysis and GST Affinity Purification

The Pharmacia Biotech pGEX-6P3 vector expression vector was used, which is an IPTG inducible expression vector. The insert is incorporated downstream of the glutathione S-transferase (GST) gene, therefore the recombinant protein is expressed as a GST chimera allowing for GST affinity chromatography using Glutathione Sepharose 4B (Amersham Biosciences) (Figure 2.2).

After induction with IPTG and incubation, the cells were harvested by centrifugation at 7741 x g (r_{max}, 8,000 rpm, Beckman JA-20 rotor) for 10 min, and resuspended in 50 mL of cold lysate buffer (PBS [140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3], 50 μM PMSF, 0.1% v/v β-mercaptoethanol, 1 mM EDTA). The cells were stored at -20°C until ready for lysis. Cells were thawed and kept on ice preceding lysis by French-press. The lysate was immediately centrifuged at 12096 x g (r_{max}, 10,000 rpm, Beckman JA-20 rotor) at 4°C for

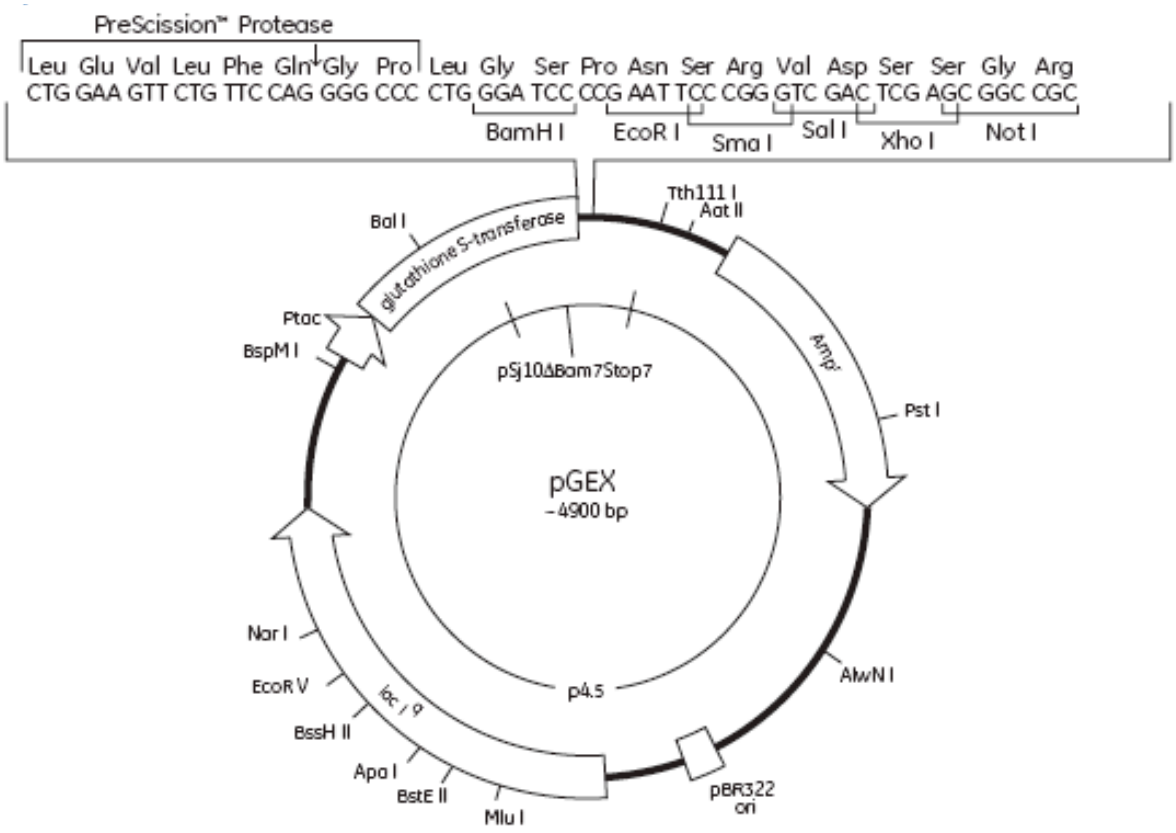


Figure 2.2: Map of the pGEX-6P3 expression vector and the multiple cloning site. Map taken from http://www6.gelifesciences.com.

45 min. The supernatant was then decanted from the pellet and loaded, at 4°C, onto a column packed with Glutathione Sepharose 4B (Amersham Biosciences) using an Äkta prime system (Amersham Biosciences). The column, stored at 4°C in 20% ethanol, was washed prior to loading with ddH_2O followed by lysate buffer. Lysate buffer was used as the running buffer. After the supernatant was loaded, the bound protein was washed with three times the volume of supernatant loaded prior to elution of the GST chimera with elution buffer (10 mM reduced glutathione, 50 mM Tris-HCl, 150 mM NaCl, pH 8). Addition of NaCl is not typically prescribed but its addition was found to solve some solubility issues. Addition of NaCl appears to broaden the elution profile of the GST chimeric protein. The eluted protein was analyzed by

SDS-PAGE. The protein was then dialyzed into 1 L of cleavage buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 8.0) prior to removal of the GST tag via PreScission protease (Amersham Biosciences) cleavage at 4°C O/N. Dialysis was carried out in two 1 L steps to remove the majority of the glutathione. After dialysis, the cleaved GST was removed by loading the protein back onto the Glutathione Sepharose 4B column, the GST binds and the cleaved protein of interest is eluted. Removal of GST was analyzed by SDS-PAGE.

2.3.14 Anion, Cation, and Size Exclusion Chromatography

2.3.14.1 Anion and Cation Exchange Chromatography

After GST affinity chromatography, the respective proteins were further purified by ion exchange chromatography. The use of cation or anion exchange was determined based upon the proteins theoretical pI, calculated from its amino acid sequence. The chromodomain was dialyzed into 2 L of its respective buffer. The buffer used for each chromodomain protein, its pI and the buffer's pH are summarized in Table 2.4. The buffer was typically 50 mM buffer, 150 mM NaCl, and 5 mM DTT. The protein was dialyzed for approximately twelve hours followed by dialysis into 2 L of running buffer containing 50 mM buffer, 5 mM DTT (Buffer A). Buffer B is identical with the exception that it contains 1 M NaCl, and is used for elution of the bound protein from the column. Using an Äkta FPLC system (Amersham Biosciences), the protein was then loaded onto the appropriate ion exchange column at a flow rate of 1 mL/min (Amersham Biosciences). The column, stored at 4°C in 20% ethanol, was washed prior to loading with three column volumes (30 mL) ddH_2O followed by three column volumes of buffer A. The bound protein was washed with three times the volume of the sample loaded with buffer A, prior to elution with a buffer B. The protein was eluted from the column with buffer B, over two column volumes, at a 50% gradient (Figure 2.3). The eluted protein was analyzed by SDS-PAGE.

2.3.14.2 Size Exclusion Chromatography (SEC)

After anion or cation exchange, the purified protein was concentrated using Millipore Amicon Ultra 5k Dalton cut-off centrifugal filter devices to approximately 1.5 mL. A 500 μL

Table 2.4: Summary of the Chromodomains Ion Exchange. Summary of the expressed chromodomain proteins pI, the type of ion exchange chromatography employed and the pH of the running buffer used.

Protein	pI	Ion Exchange	Buffer A pH	
D. melanogaster				
MOF(371-440)	5.84	cation	Sodium Acetate	5.0
		anion	Tris·HCl	8.0
MSL3 (2-96)	8.97	cation	Sodium Acetate	5.5
MSL3 (2-91)	8.60	cation	Sodium Acetate	5.5
H. sapiens				
MOF (20-93)	5.31	anion	Tris·HCl	8.0
MSL3 (2-93)	9.72	cation	Bis-Tris-propane	6.5
S. pombe				
SWI6 (77-135)	4.52	anion	Tris.HCl	8.0

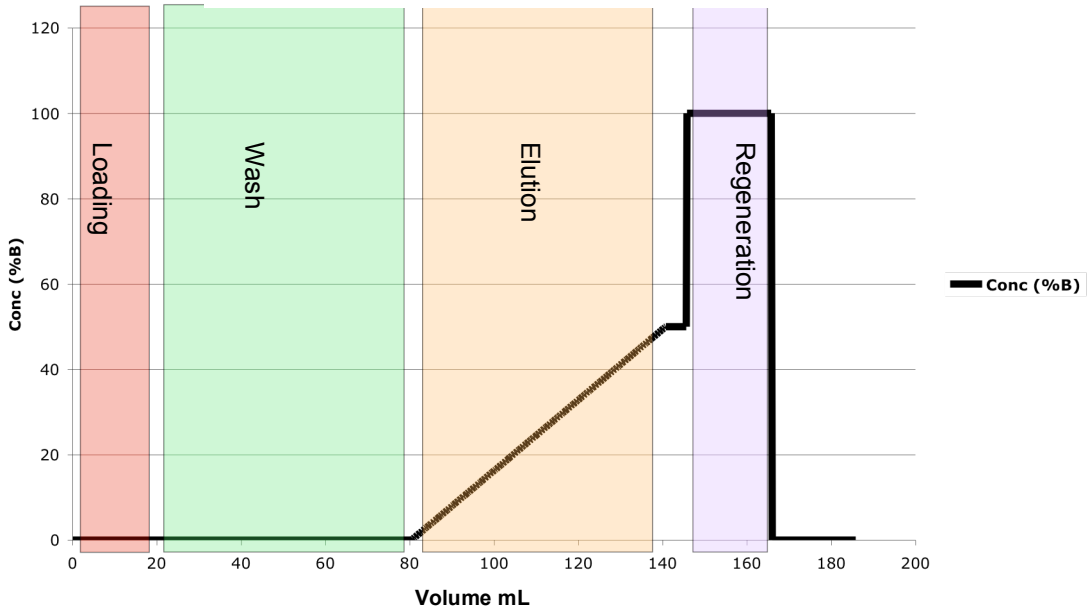


Figure 2.3: Ion exchange chromatography profile. A typical profile used for anion or cation exchange chromatography showing the loading (red), washing (green), elution gradient (orange), and regeneration (purple) stages.

aliquot was then loaded onto a Superdex 75 10/300 GL (Amersham Biosciences) size exclusion column, pre-equilibrated in PBS containing 5 mM DTT, at a flow rate of 0.5 mL/min. Care was taken to not let the pressure exceed 1 MPa. The eluted peak was analyzed by SDS-PAGE.

2.3.15 Isotopic labeling of the *D. melanogaster* MOF and MSL3 Chromodomains

To investigate the structure of the *D. melanogaster* MOF and MSL3 chromodomains by nuclear magnetic resonance ^{15}N and ^{13}C isotopically labeled protein was prepared (Marley *et al*., 2001). A 5 mL aliquot of LB containing 100 μg/mL ampicillin, inoculated with a single *E. coli* BL21 colony transformed with approximately 20 ng of either the recombinant MOF or MSL3 plasmid, was incubated at 37°C O/N. Two liters of LB containing 100 μg/mL ampicillin, was inoculated with 1 mL of the O/N and incubated with shaking at 37°C until an $OD_{(600nm)}$ was reached. Under sterile conditions, the cells were harvested by centrifugation at 7741 x g (r_{max}, 8,000 rpm, Beckman JA-20 rotor) for 10 min, and washed (3 x 50 mL) with warm (37°C) M9 minimal medium (22 mM KH_2PO_4, 90 mM Na_2HPO_4, 8 mM NaCl, 22 mM D-glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$). After washing, the cells were resuspended and added to 1 L M9 minimal medium containing either 1 g/L $^{15}NH_4Cl$ or 1 g/L ^{13}C-D-Glucose. After incubating with shaking at 37°C for 20-30 mins, the temperature was reduced to 25°C, the cells were induced with 0.1 mM IPTG and incubated O/N. The cells were subsequently harvested and lysed, as described above (section 2.3.13). The isotopically labeled protein was purified by ion exchange and SEC, as described above (2.3.14). Dr. Hideo Iwai (University of Saskatchewan) measured the ^{1}H-^{15}N and ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) Spectra at the Saskatchewan Structural Sciences Centre (SSSC) on the Bruker 500 MHz Avance NMR spectrometer.

2.3.16 Site-Directed Mutagenesis

2.3.16.1 Site-Directed Mutagenesis of the *D. melanogaster* MOF Chromodomain

The absence of a Met residue in the *D. melanogaster* MOF chromodomains meant that one had to be engineered into the protein before a seleno-methionine (Se-Met) derivative could

be prepared. A total of nine mutants were made, summarized in Table 2.5, using the Stratagene QuickChange II Site-Directed Mutagenesis Kit and following the instructions contained therein. Mutant plasmid was prepared as described above (section 2.3.11) and the mutation was confirmed by DNA sequencing.

Table 2.5: Summary of the *D. melanogaster* MOF chromodomain methionine mutations. The mutated codon is shown in **bold**. The forward primer is listed first, the reverse primer second.

Mutant		DNA Sequence
Ile385Met	5'	CCC GAC AAG ATC TAT TTT **ATG** CGG CGC GAG GAT GGC ACC GTG
	5'	CAC GGT GCC ATC CTC GCG CCG **CAT** AAA ATA GAT CTT GTC GGG
Ile432Met	5'	GGA TGG GTG GGC AGG CAT AGG **ATG** TCG GAC AAT GCG GAT GAC TTG GGC
	5'	GCC CAA GTC ATC CGC ATT GTC CGA **CAT** CCT ATG CCT GCC CAC CCA TCC
L398Met	5'	CGT GGC CAG GTG **ATG** CAG TCT CGC ACC ACT GAG AAT GCG
	5'	CGC ATT CTC AGT GGT GCG AGA CTG **CAT** CAC CTG GCC ACG
Leu419Met	5'	C GTC CAC TAT GTT GGT **ATG** AAT CGC CGC CTG GAC GG
	5'	CC GTC CAG GCG GCG ATT **CAT** ACC AAC ATA GTG GAC G
Leu423M	5'	GGT CTC AAT CGC CGC **ATG** GAC GGA TGG GTG GGC AGG
	5'	CCT GCC CAC CCA TCC GTC **CAT** GCG GCG ATT GAG ACC
Val397Met	5'	CC GTG CAC CGT GGC CAG **ATG** CTT CAG TCT CGC
	'	GCG AGA CTG AAG **CAT** CTG GCC ACG GTG CAC GG
Val414Met	5'	GCC GCT CCC GAT GAG TAC TAC **ATG** CAC TAT GTT GG
	5'	CC AAC ATA GTG **CAT** GTA GTA CTC ATC GGG AGC GGC
Val427Met	5'	CGC CTG GAC GGA TGG **ATG** GGC AGG CAT AGG
	5'	CCT ATG CCT GCC **CAT** CCA TCC GTC CAG GCG
Ile327Met	5'	CCC CTG GGA TCC CAA AAG **ATG** GAT ATA AGC GAA AAT CCC
	5'	GGG ATT TTC GCT TAT ATC **CAT** CTT TTG GGA TCC CAG GGG

2.3.16.2 Site-Directed Mutagenesis of the *D. melanogaster* MSL3 Chromodomain

During ion exchange and SEC, it was observed that the *D. melanogaster* MSL3 chromodomain forms oxidized cysteine oligomers (discussed in greater detail in section 3.2.1). A reducing agent was added at every stage of the purification right from lysis. DTT was typically used, though due to its relative instability (it spontaneously oxidizes to form a six-member ring with an internal disulfide bond) it is not suitable for use in crystallization screens due to the relatively long period of time required for protein nucleation and crystallization. To maintain the *D. melanogaster* MSL3 chromodomain (2-91) in its reduced monomeric form, the reducing agent Tris(2-carboxyethyl)phosphine HCl (TCEP hydrochloride) was added to the crystallization buffer. Despite TCEP's relative stability, it still does not ensure that the protein is not oxidizing, which in theory would be an obstacle to crystallization as the protein is no longer homogeneous. To address this issue, mutation of Cys66, one of the two cysteine residues, was investigated. The other Cys residue (Cys19) is part of the hydrophobic core, and is therefore unlikely to be solvent exposed. Primers were designed for mutagenesis of the Cys residue to Ser, due to the conserved nature of the mutation, using the Stratagene QuickChange II Site-Directed Mutagenesis Kit (Table 2.6).

Table 2.6: Primers used in the site-directed mutagenesis of C66 to Serine using the Stratagene QuickChange II Site-Directed Mutagenesis Kit. The mutated codon is shown in **bold**. The forward primer is listed first, the reverse primer second.

Mutant		DNA Sequence
C66S	5'	CGT CCA TCC TAC GAT CGC **AGT** GTG CGC GCC ACC
	5'	GGT GGC GCG CAC **ACT** GCG ATC GTA GGA TGG ACG

2.3.17 Incorporation of Se-Met into Met Mutant *D. melanogaster* MOF

A 5 mL aliquot of LB containing 100 μg/mL ampicillin, inoculated with a single BL21 colony transformed with the desired Met mutant MOF plasmid, was incubated at 37°C O/N. Two liters of LB containing 100 μg/mL ampicillin was inoculated with 1 mL of the O/N and incubated with shaking at 37°C until an $OD_{(600nm)}$ of 0.6 was reached. Under sterile conditions the cells were harvested by centrifugation at 7741 x g (r_{max}, 8,000 rpm, Beckman JA-20 rotor) for 10 min, and washed (3 x 50 mL) with warm (37°C) M9 minimal medium. The cells were then washed a final time with 50 mL of M9 minimal medium containing amino acids (100 mg/L L-Lys, 100 mg/L L-Thr, 100 mg/L L-Phe, 50 mg/L L-Val, 50 mg/L L-Leu, 50 mg/L L-Ile) to inhibit endogenous Met synthesis. After washing, the cells were resuspended and added to 1 L M9 minimal medium containing the above amino acids and 60 mg/L L-seleno-methionine (Se-Met). After incubating with shaking at 37°C for 20-30 mins, the temperature was reduced to 25°C and the cells were induced with 0.1 mM IPTG and incubated O/N. The cells were subsequently harvested and lysed as described above (section 2.3.13). The Se-Met MOF was purified by anion exchange and SEC as described above (section 2.3.14).

2.3.18 Limited Proteolysis

Approximately 0.1 mg of trypsin was dissolved in 1 mL of 10% glycerol and PBS and its molar concentration determined by $UV_{(280nm)}$ spectroscopy. Trypsin was added to the recombinant protein of interest at a molar ratio of 1:1000. The protease digestion was allowed to proceed at room temperature over an hour with 5 μL samples taken at 5, 10, 20, 40, and 60 min. The reaction was stopped via the immediate addition of 1 μL of 200 μM PMSF and 5 μL SDS loading dye prior to boiling for 1 min (Lane *et al*., 2006). The protease-digested samples were then analyzed by SDS-PAGE.

2.3.19 Vapour Diffusion Crystallization Screens

The crystallization of a macromolecule by vapour diffusion involves the suspension of a droplet containing protein, buffer and precipitant over a reservoir containing the precipitant at a higher concentration. The difference in precipitant concentration causes diffusion of water vapor from the droplet into the reservoir, which ideally results in the protein concentration in

the droplet becoming super-saturated and the protein crystallizing. The protein drop was suspended on a plastic microscope coverslip and suspended over a single well of a 24 well tissue culture tray. The well was sealed with vacuum grease.

Initial crystallization screens of the *D. melanogaster* MOF and MSL3, *H. sapiens* MOF, and *Schizosaccharomyces pombe* SWI6 chromodomains used the Emerald Biosciences Wizard I and II and Hampton Research Crystal Screen I and II commercial sparse matrix screens.

D. melanogaster MOF chromodomain crystals were initially identified growing in condition 16 of the Emerald Biosciences Wizard II screen, which contained 1 M sodium citrate and 100 mM CHES pH 9.5. Citrate concentration and pH were subsequently screened to optimize crystal size. Optimal conditions for the growth of large crystals was around 1 M sodium citrate, 10% glycerol (v/v), and 100 mM CHES, pH 9.2.

2.3.20 Iodide and Bromide Crystal Soaking

In an attempt to incorporate an anomalous scatterer (iodide [I^-] or bromide [Br^-]) into the crystal lattice to attain an anomalous signal, native crystals were soaked in a cryo-solution containing 500 mM NaBr or 500 mM KI, 1.5 M sodium citrate, 100 mM CHES, pH 9.5 for 20-40 sec. The increased sodium citrate concentration and pH, compared to the crystallization conditions stated above, is required to prevent the crystals from dissolving. The crystal was then immediately mounted on the goniostat and frozen in a stream of nitrogen at 114 K.

2.3.21 X-ray Diffraction Experiments

Initial diffraction experiments of the native recombinant *D. melanogaster* MOF chromodomain (371-440) crystals and iodide-soaked crystals were carried out at the Saskatchewan Structural Sciences Centre using a DX8 Proteum rotating Copper (Cu) anode diffractometer with a Kappa goniostat and Proteum 4K charge-coupled device (CCD) detector. Crystals were diffracted under a stream of nitrogen at 114 K using a Kryoflex low temperature apparatus. The diffraction images were processed using SaintPlus and ProScale (Bruker AXS Inc), and HKL2000 (Otwinowski and Minor, 1997).

A native data set was kindly collected by Prof. Louis Delbaere's Laboratory (University of Saskatchewan) at the Advanced Photon Source on the BioCARS 14 BM-C beamline, using a ADSQ Quantum 315 detector. Crystals were diffracted under a stream of nitrogen at 105 K.

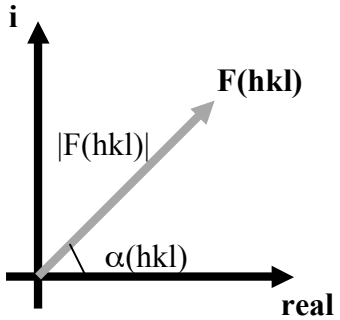
The diffraction images were processed using the HKL2000 software package (Otwinowski and Minor, 1997).

Diffraction of Se-Met *D. melanogaster* MOF L423M mutant crystals, native crystals, and native crystals soaked in an iodide or bromide cryo-solution was carried out at the Canadian Light Source (CLS) on the Canadian Macromolecular Facility 1 (CMCF1) 08ID-1 beamline using a MAR225 Detector. The diffraction images were processed using the HKL2000 (Otwinowski and Minor, 1997) software package.

2.3.22 Phase Determination

A photon has both amplitude, which is proportional to the square root of the intensity, and phase. In a diffraction experiment, only the intensity of the diffracted photon is measured since the relative phase information is lost. Determination of the phases by either experimental or theoretical methods is essential to the solution of an X-ray crystal structure. Central to solution of the phase problem is the Patterson map (Figure 2.4).

In crystallography the electron density map is calculated via an inverse Fourier transform of the structure factors (amplitude and phase). The Patterson map is the Fourier transform of the squared amplitudes. Peaks in the Patterson map represent vectors between atoms. There are three types of peaks in the Patterson map arising from intramolecular vectors, intermolecular vectors, and self-vectors. A vector has both direction and magnitude (or length).



$$\rho_{(xyz)} = 1/V \sum \sum_h \sum_k |F_{(hkl)}| e^{-2\pi i(hx+ky+lz) + i\alpha(hkl)}$$

Figure 2.4: Argand diagram depicting a structure factor, its phase (α[hkl]) and its magnitude (|F[hkl]|). The equation for the calculation of the electron density ($\rho_{[xyz]}$) is the sum of the structure factor amplitudes and phases.

Self-vectors are vectors between an atom and itself and therefore have no magnitude, therefore giving rise to a large peak at the origin of the Patterson map. Intramolecular vectors will be shorter than intermolecular vectors and will therefore give rise to a peak closer to the origin of the Patterson map. The height of a peak in the Patterson map is directly proportional to the number of electrons of a given atom, therefore the heavier the atom the larger the peak in the Patterson map. The original position of the atoms can be directly calculated from the Patterson map for small molecule crystals. Solution of the Patterson map becomes increasingly difficult as the number of atoms increases, which directly correlates to the number of peaks within the Patterson map increasing. The large number of atoms in a protein molecule makes *ab initio* solution of the Patterson map practically impossible.

2.3.22.1 Molecular Replacement: PHASER

The molecular replacement software, PHASER (McCoy *et al*., 2007), was employed initially to determine phases for the diffraction amplitudes of the *D. melanogaster* MOF chromodomain crystals, using the solution NMR structure published during the course of this research (Nielsen *et al*., 2005)(PDB 2bud). The 2bud.pdb file is an ensemble of 25 conformers of residues 367-454. Based upon superposition of the ensembles members, the search model was truncated to residues 378-443 to remove disordered or heterogenous regions. Single conformers along with the entire ensemble were used in separate searches with and without the side-chains truncated to the β-carbon atom.

To calculate phases from the model, the model needs to be correctly positioned into the crystal unit cell to approximate the structure that we are trying to solve. Three rotation angles and three translational distances define the models position. The intramolecular vectors of the Patterson map are defined by the orientation of the molecule only and not its position in the cell. Therefore the model can be oriented properly by using data from the Patterson map close to the origin. Once a rotation solution(s) has been found, the intermolecular data from the Patterson map is used to translate the model into its correct position in the unit cell.

PHASER sequentially carries out an anisotropy correction of the diffraction data, cell content analysis, a rotation function to orientate, and a translation function to position the model in the unit cell, using maximum likelihood methods. Molecular replacement basically measures the correlation between the Patterson function calculated from the diffraction data and

the Patterson function calculated from the model. The likelihood function is a probability measuring the agreement between the two. A high probability indicates that the model's position and orientation within the unit cell closely approximates that of the structure to be resolved and phases from the model can be then calculated and used to calculate an initial electron density map.

2.3.22.2 Single Isomorphous Replacement (SIR)

Isomorphous replacement involves the soaking or co-crystallization of a heavy atom into the crystal. Data is collected separately on a native crystal and the isomorphous heavy atom derivative crystal. The difference in the structure factor amplitudes between the two data sets is due to the contribution of the heavy atom(s). The structure factor of the heavy atom derivative (F_{PH}) is the vector sum of the heavy atom structure factor (F_H) and the native protein atoms structure factor (F_P). Therefore a differences Patterson map can be calculated. The peaks in the calculated Patterson map will theoretically only represent a vector between two heavy atoms. As only a few heavy atoms are incorporated, the differences Patterson map can be solved directly. SIR gives two trigonometrically possible values for the phase. In theory, multiple isomorphous replacement (MIR) (i.e. more than one derivative data sets) gives only one possible phase. In reality, compound errors and approximations give rise to a phase probability distribution for each measured reflection. Phase information, calculated from the position of the heavy atoms in the unit cell, can then be used to calculate an initial electron density map.

2.3.22.3 Multi-wavelength Anomalous Dispersion (MAD)

Friedels Law states that a reflection from a plane with Miller indices (h,k,l) will have identical magnitude but opposite phase as a reflection from a plane with Miller indices (-h,-k,-l). Friedels Law is obeyed only when there is no anomalous dispersion. Anomalous dispersion is a phenomenon where an atom's electrons absorb a photon of a specific wavelength and reemits it after a delay, with a phase shift. The anomalous scattering effect is generally more significant the heavier the atom, but also depends strongly on the exciting wavelength. The energy of a photon (E) is inversely proportional to its wavelength (λ), c is the speed of light in a vacuum (3×10^8 m/s) and h is Plancks constant (6.626×10^{-34} J·s).

$$E=(hc)/\lambda$$

The energy of the photon that a specific element absorbs is known as that element's absorption edge. In a typical MAD experiment, diffraction data is collected at three wavelengths; at the peak of the absorption, at the inflection point of the absorption (when the dispersive component is at a minimum), and at a remote wavelength where there is minimal anomalous dispersion. Analysis of the anomalous differences Patterson map, calculated using the three data sets, allows for the calculation of the anomalous scatterers position and the subsequent calculation of phases, analogous to MIR.

Multi-wavelength anomalous dispersion data from Se-Met mutant MOF crystals were collected at the CLS on the CMCF-1 08ID-1 beamline. A fluorescence scan was run initially to confirm selenium incorporation. Data was collected at the peak (12.6558 keV) and inflection (12.6553) points of the selenium absorption edge. Peak and inflection were chosen at the suggestion of the beamline scientist, although if only two wavelengths are chosen, peak and remote are typically used. Initial data was collected over 120° with 0.5° oscillations over 2 sec. Beam attenuation, oscillation, and exposure time were refined subsequently to reduce radiation damage.

Single isomorphous replacement with anomalous signal (SIRAS) data is collected at a single wavelength, which optimally reduces exposure time and hence radiation damage to the data. It is similar to SIR but uses the anomalous signal to help resolve the phase ambiguity.

3.0 Results

3.1 Purification and Isotopic Labeling of the *D. melanogaster* MSL3 Chromodomain (2-96)

The original *D. melanogaster* MSL3 chromodomain (2-96) construct was designed by Dr. S. Moore and cloned into the pGEX-6P3 expression vector by B. McClellan (Massey University). The theoretical pI and molecular weight (MW) of the *D. melanogaster* MSL3 chromodomain (2-96), based upon the amino acid sequence is 8.97 and 11398 Da, respectively.

When the *D. melanogaster* MSL3 chromodomain (2-96) GST chimera was analyzed by SDS-PAGE (17% acrylamide:bis [37.5:1]) after GST affinity chromatography, a large single band was observed at approximately 36 kDa, indicating that the construct expressed well, and that the purification step was successful in removing the majority of impurities. Remaining impurities are observed as several faint bands on the SDS-PAGE gel. The protein appeared to be relatively stable as no visible degradation products, such as GST, were observed (Figure 3.1 A). When digested with PreScission protease, two fragments were observed on the SDS-PAGE gel, as expected, with molecular weights consistent with the MSL3 chromodomain (11 kDa) and GST (25 kDa) (Figure 3.1 B). After removal of the cleaved GST and further purification by cation exchange chromatography, the *D. melanogaster* MSL3 chromodomain (2-96) was visibly homogenous, based upon SDS-PAGE (Figure 3.1 C).

The *D. melanogaster* MSL3 chromodomain was isotopically labeled with ^{15}N and ^{13}C as outlined in section 2.3.15. Dr. Hideo Iwai collected the ^{1}H-^{15}N and ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) spectra at the Saskatchewan Structural Sciences Centre (SSSC) on a Bruker 500 MHz Avance NMR spectrometer (Figure 3.2). The observation and assignment of multiple peaks of weak intensity in the ^{1}H-^{15}N HSQC spectra indicated that the last five residues of the C-terminus of the *D. melanogaster* MSL3 chromodomain (2-96) were undergoing rapid proton exchange, indicative of a disordered structure. The disordered section could prevent a solution of the structure by NMR and/or prevent the protein from crystallizing. Dr. S. Moore redesigned the domain, truncating the last five C-terminal residues (Figure 3.3).

The NMR studies were not pursued further after Dr. Hideo Iwai left the University of Saskat

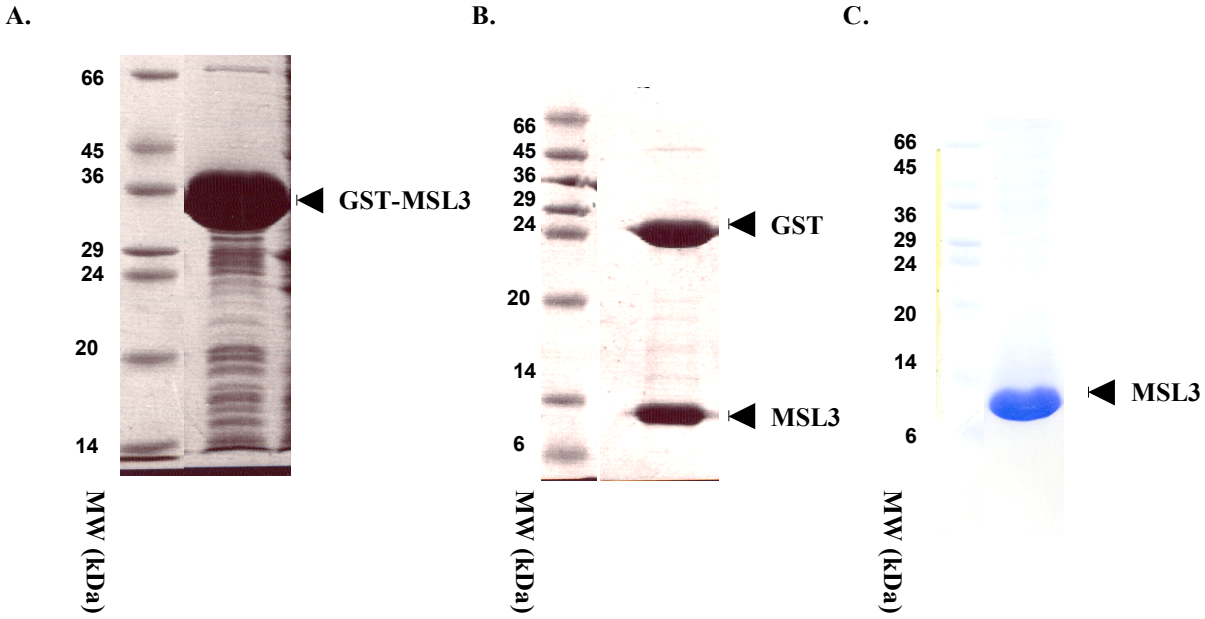


Figure 3.1: Purification of the *D. melanogaster* MSL3 chromodomain (2-96). **A.** SDS-PAGE (17% acrylamide:bis [37.5:1]) of the GST *D. melanogaster* MSL3 chromodomain (2-96) purified by GST affinity chromatography using a glutathione Sepharose 4B column. Molecular weight (MW) markers are shown. **B.** SDS-PAGE of the *D. melanogaster* MSL3 chromodomain (2-96) after cleavage with PreScission protease. Molecular weight (MW) markers are shown. **C.** SDS-PAGE of the *D. melanogaster* MSL3 chromodomain (2-96) after cation exchange chromatography. Molecular weight (MW) markers are shown.

3.2 *D. melanogaster* MSL3 Chromodomain (2-91)

3.2.1 Purification of the *D. melanogaster* MSL3 Chromodomain (2-91)

The truncated *D. melanogaster* MSL3 chromodomain (2-91) construct, designed by Dr. S. Moore based upon the ^{1}H-^{15}N HSQC findings, was cloned into the pGEX-6P3 expression vector. Ewa Kerc (University of Saskatchewan) contributed to the cloning, ligation, transformation, and plasmid purification. The theoretical pI and molecular weight (MW) of the *D. melanogaster* MSL3 chromodomain (2-91) based upon the amino acid sequence is 8.6 and 10830 Da, respectively. SDS-PAGE (17% acrylamide:bis [37.5:1]) of the eluate after GST

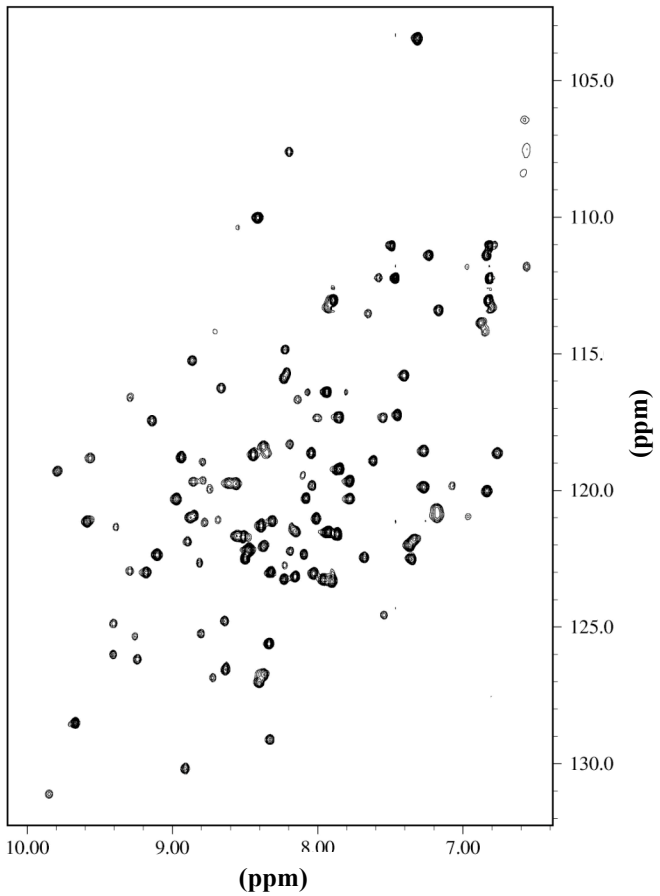


Figure 3.2: ^{1}H-^{15}N HSQC spectrum of the *D. melanogaster* MSL3 chromodomain.

```
             1           11          21          31          41
             |           |           |           |           |
MSL3 (2-96) MTELRDETPL FHKGEIVLCY EPDKSKARVL YTSKVLNVFE RRNEHGLRFY
MSL3 (2-91) MTELRDETPL FHKGEIVLCY EPDKSKARVL YTSKVLNVFE RRNEHGLRFY

             51          61          71          81          91
             |           |           |           |           |
MSL3 (2-96) EYKIHFQGWR PSYDRCVRAT VLLKDTEENR QLQRELAEAA KLQIRG
MSL3 (2-91) EYKIHFQGWR PSYDRCVRAT VLLKDTEENR QLQRELAEAA K
```

Figure 3.3: Sequence alignment of the two *D. melanogaster* MSL3 chromodomain constructs designed by Dr. S. Moore. The two cysteine residues in the construct are highlighted and underlined.

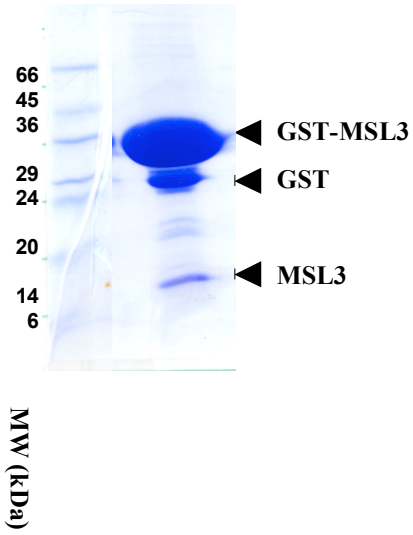
affinity purification revealed a large band running at approximately 36 kDa, which was consistent with the size of the GST-MSL3 (Figure 3.4 A). The bands observed at approximately 25 kDa and 14 kDa, which were likely to be GST and MSL3, respectively, indicate some proteolysis had occurred. Other than the degradation products, no other significant bands were observed on the SDS-PAGE gel, indicating that the GST affinity purification was successful in purifying the protein from the lysate. When digested with PreScission protease, two fragments were observed on the SDS-PAGE gel, as expected, with molecular weights consistent with the MSL3 (2-91) chromodomain (approximately 10 kDa) and GST (25 kDa) (Figure 3.4 B). Although the MSL3 (2-91) protein is smaller than the MSL3 (2-96) protein, the difference in molecular weight was too small to distinguish by SDS-PAGE or SEC. DNA sequencing previously confirmed the correct size of the constructs.

After GST affinity purification and PreScission protease digestion, the MSL3 (2-91) chromodomain was purified to homogeneity by cation exchange and size exclusion chromatography (SEC) (Figure 3.4 C-F). After size exclusion chromatography, the MSL3 (2-91) chromodomain was concentrated to approximately 20 mg/mL in a total volume of 350 μL, a total yield of approximately 7 mg of protein from 2 L of culture. The concentration was calculated using Beer's Law (Equation 3.1) and the solutions absorbance, measured at 280 nm. SDS-PAGE indicated that the MSL3 (2-91) chromodomain was homogenous. The observation of a single peak on the SEC75 chromatogram indicated that the MSL3 (2-91) chromodomain was present as a homogenous species (Figure 3.4 E). By measuring the retention volumes of proteins of known molecular weight (Table 3.1) a standard curve can be constructed using Equation 3.2, from which the experimental molecular weight of an unknown protein can be approximated (Figure 3.5).

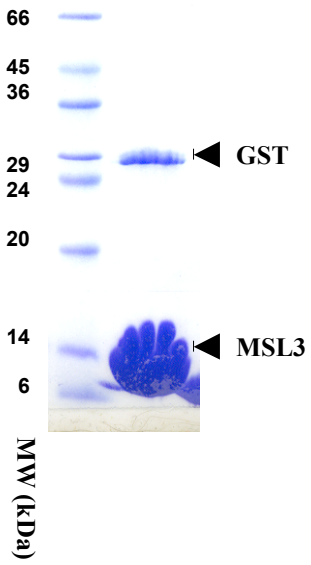
$$\mathbf{A = \varepsilon lc}$$

Equation 3.1: Beer's Law. **A** is absorbance (no units), ε is the molar absorbtivity (L/mol/cm)**, l** is the path length of the sample (cm) and **c** is the concentration of the absorbing species in solution (mol/L).

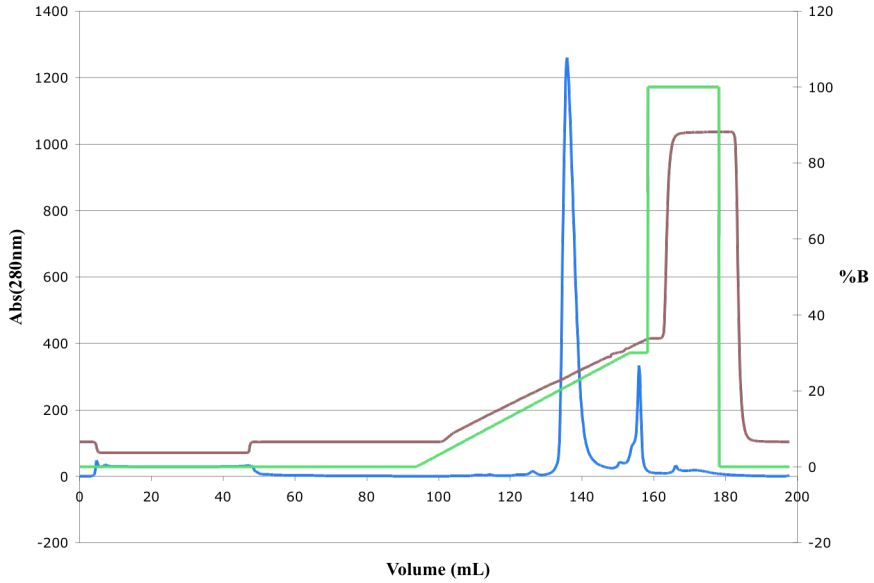
A.


66
45
36
29
24
20
14
6
GST-MSL3
GST
MSL3
MW (kDa)

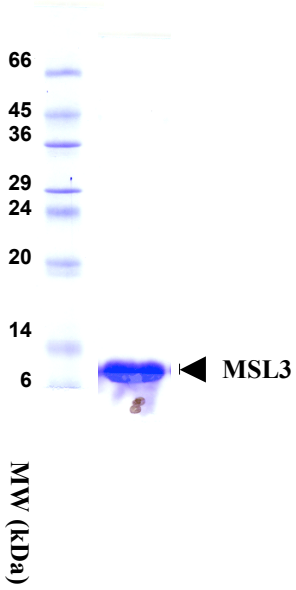
B.


66
45
36
29
24
20
14
6
GST
MSL3
MW (kDa)

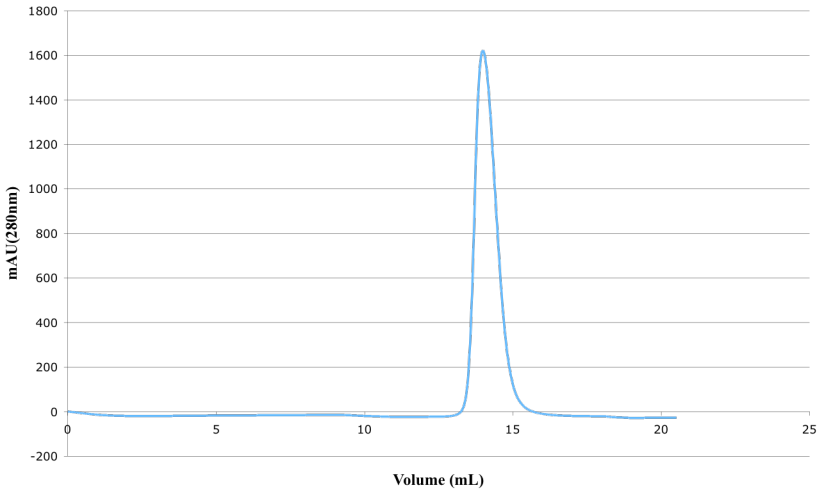
C.


1400
1200
1000
800
600
400
200
0
-200
120
100
80
60
40
20
0
-20
0
20
40
60
80
100
120
140
160
180
200
Abs(280nm)
%B
Volume (mL)

D.


66
45
36
29
24
20
14
6
MSL3
MW (kDa)

E.



F.

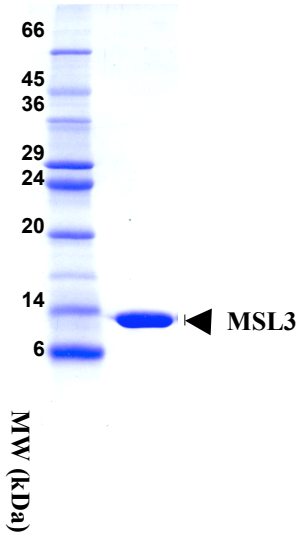


Figure 3.4: Purification of the *D. melanogaster* MSL3 chromodomain (2-91). **A.** SDS-PAGE (17% acrylamide:bis [37.5:1]) of the eluted GST-MSL3 after GST affinity purification. Molecular weight markers are shown. **B.** SDS-PAGE of the GST-MSL3 after cleavage with PreScission protease and removal of the GST by a subsequent round of GST affinity purification after the glutathione had been removed by dialysis. **C.** Cation exchange chromatogram of the MSL3 chromodomain. The cleaved MSL3 chromodomain was loaded onto a SourceS cation exchange column at 1 mL/min. The protein was bound and washed with buffer A (see section 2.3.13.1) prior to a gradient elution with buffer B. Absorbance (280 nm) is shown in blue (mAU is absorbing units x 1000), conductivity (indicating salt concentration) in brown (mS/cm), and concentration (percentage of buffer B) is shown in green (%B). **D.** SDS-PAGE of the MSL3 chromodomain after cation exchange purification. **E.** Size exclusion chromatography of the cation exchange fractions, determined by SDS-PAGE to contain the MSL3 chromodomain, after pooling and concentration. 500 µL was loaded onto a Superdex 75 column at a flow rate of 0.5 mL/min. The eluate was collected in 1 mL fractions. **F.** SDS-PAGE of the peak that eluted from the SEC75 column.

$$K_{av} = \frac{V_e - V_o}{V_t - V_o}$$

Equation 3.2: Definition of K_{av}. K_{av} is a constant for a specific protein of specific molecular weight. V_e is the elution volume of the protein, V_o the column void volume and V_t the total bed volume.

Table 3.1: Standard curve data for construction of the Superdex 75 molecular weight standard curve. Data provided by Dr. M. Lane (University of Saskatchewan).

Standard	Ve	Vo	Vt	Kav	MW	logMW
Albumin	9.84	7.84	24	0.12	67000	4.826
Ovalbumin	10.88	7.84	24	0.19	43000	4.6335
Chymotrypsinogen A	12.8	7.84	24	0.31	25000	4.3979
Ribonuclease A	13.8	7.84	24	0.37	13700	4.1367
Aprotinin	15.73	7.84	24	0.49	6500	3.8129
Vitamin B12	19.15	7.84	24	0.7	1355	3.1319

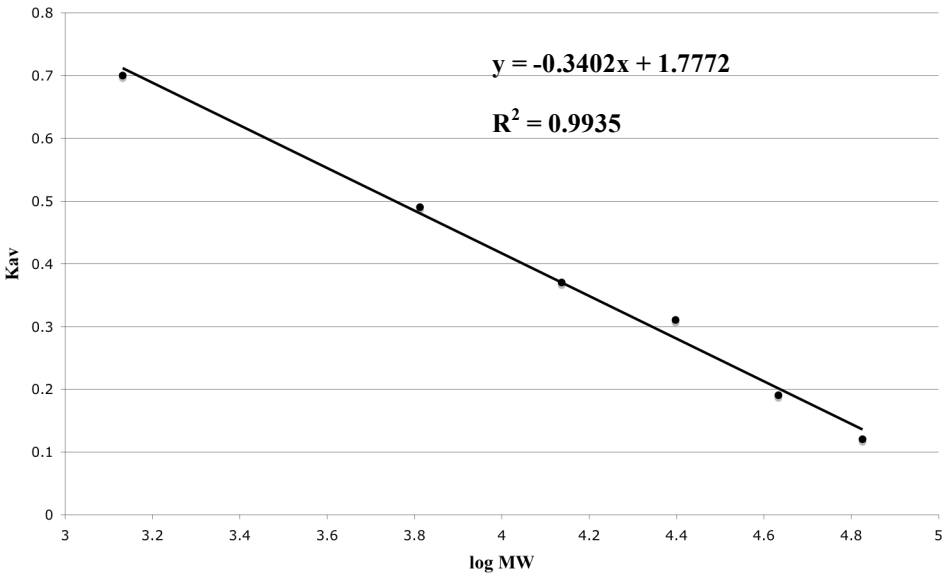


Figure 3.5: Superdex 75 molecular weight calibration curve. Plot of K_{av} versus the log molecular weight of the standards. The slope of the best-fit is shown along with the regression score. The data was provided by Dr. M. Lane, formerly of the University of Saskatchewan.

The theoretical molecular weight of the *D. melanogaster* MSL3 (2-91) chromodomain is 10830.3 Da. From the SEC75 chromatogram, the MSL3 (2-91) chromodomain had a retention volume of 13.99 mL, which correlates to a molecular weight of 12691 Da. The *D. melanogaster* MSL3 (2-91) chromodomain is therefore likely to be homogenous in solution as a monomer. During the process of refining the purification process, it was observed that a reducing agent had to be present throughout, beginning at lysis. If a reducing agent was not present, oxidized MSL3 (2-91) oligomers were observed at the size exclusion stage, with retention volumes of 10.75 mL and 5.61 mL, indicative of their higher molecular weight than the monomer (Figure 3.6).

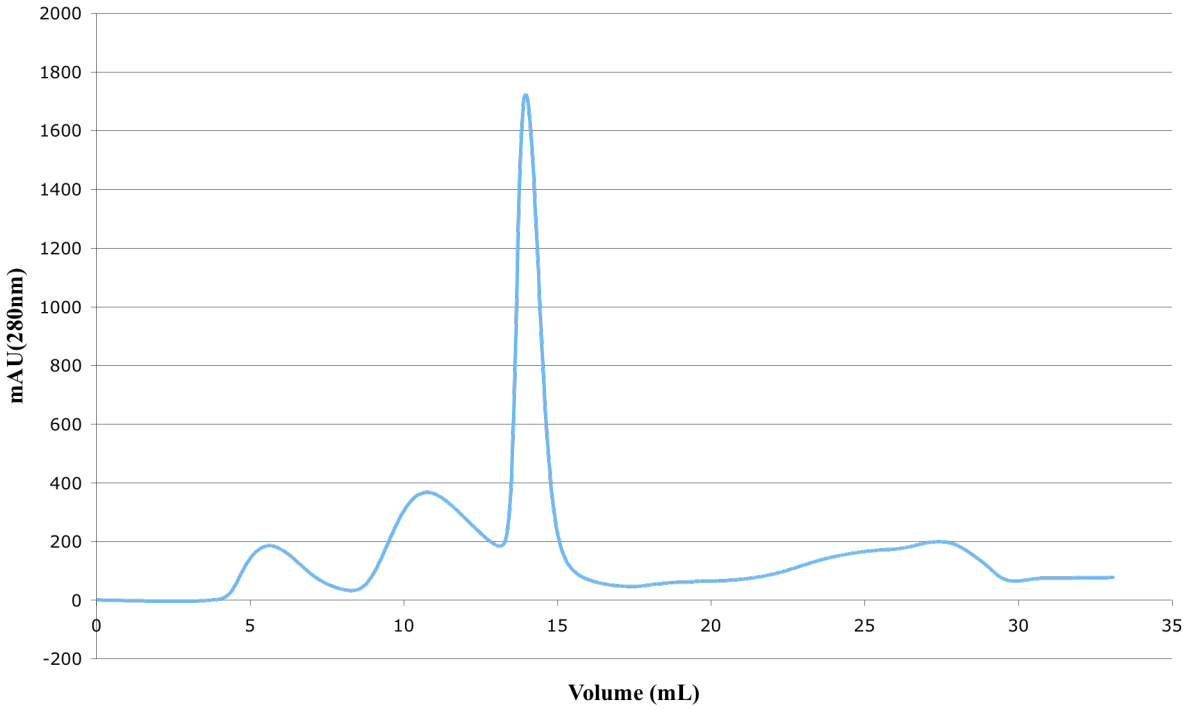


Figure 3.6: Size exclusion chromatography of oxidized *D. melanogaster* MSL3 (2-91) chromodomain. 500 µL was loaded onto a Superdex 75 column at a flow rate of 0.5 mL/min. Absorbance (280 nm) is shown in blue (mAU is absorbing units x 1000).

3.2.2 Limited Proteolysis

In order to investigate the stability of the truncated *D. melanogaster* MSL3 chromodomain (2-91), the purified protein was subjected to limited trypsin proteolysis (Figure 3.7). At a molar ratio of 1:1000 trypsin to MSL3 (2-91) chromodomain, little proteolytic degradation was observed after an hour at room temperature. This indicated that the expressed domain was likely to be folded and stable. Significant proteolytic degradation was observed at a molar ratio of 1:100 trypsin to MSL3 (2-91) chromodomain. A molar ratio of 1:100 was chosen to demonstrate that the protease was active, as significant proteolysis would be expected at this ratio.

3.2.3 Oligonucleotide Association with the *D. melanogaster* MSL3 Chromodomain

During the GST affinity purification, it was observed that the eluted fraction absorbed strongly at 260 nm when analyzed by UV spectroscopy, indicating the likely presence of an oligonucleotide (Figure 3.8 A). The species was able to be ethanol precipitated, stained with ethidium bromide when run on an agarose gel, and digested by RNAse A and Benzonase (an oligonuclease) (Figure 3.8 B). The strong absorption of the oligonucleotide at 260 nm meant that its presence prevented determination of the protein concentration by UV spectroscopy at 280 nm, and prevented monitoring the elution of the protein from the ion exchange column by UV spectroscopy at 280 nm. To avoid these complications, benzonase was added to the lysate in subsequent purifications.

3.2.4 Vapour Diffusion Crystallization Screens

After ion exchange and SEC chromatography, the purified *D. melanogaster* MSL3 (2-91) chromodomain was concentrated to 20 mg/mL and equilibrated into a buffer containing 20 mM Bis-Tris-propane, pH 7, 50 mM NaCl, 5 mM TCEP. Crystallization conditions were initially investigated using the Emerald Biosciences Wizard I and II and Hampton Research Crystal Screen I and II commercial sparse matrix screens. Duplicate drops were set up, with protein alone and protein with a histone H4 N-terminal peptide at a 20 % molar excess. Both an unmodified H4 N-terminal peptide and an H4K16Ac peptide were used independently. No conditions were identified that yielded observable crystals.

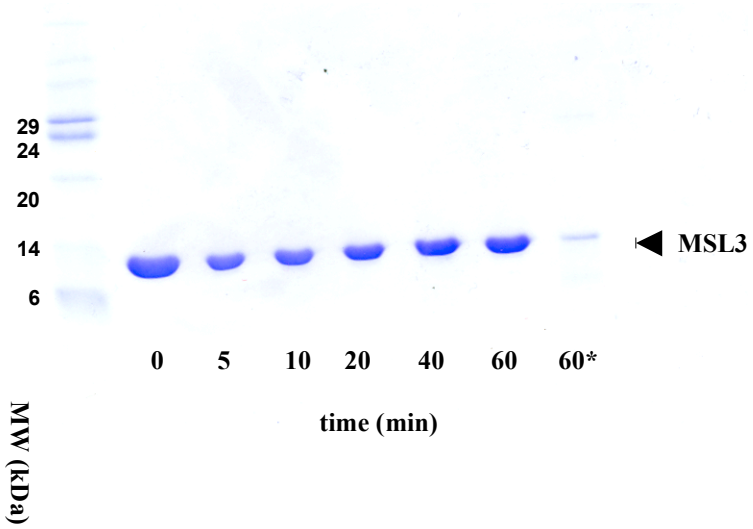
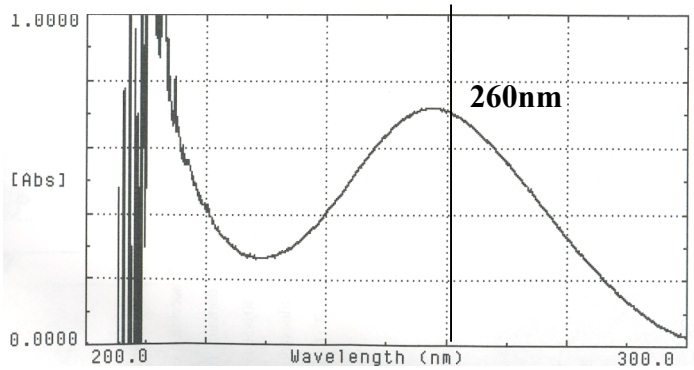


Figure 3.7: Limited proteolysis of the *D. melanogaster* MSL3 chromodomain (2-91). SDS-PAGE (17% acrylamide:bis [37.5:1]) of the limited proteolysis of the *D. melanogaster* MSL3 chromodomain (2-91). 5 µL of 20 mg/mL MSL3 (2-91) was incubated with trypsin at a 1:1000 molar ratio for 5, 10, 20, 40 and 60 minutes. A sample was also incubated at a molar ratio of 1:100 for 60 mins (60*).

A.



B.

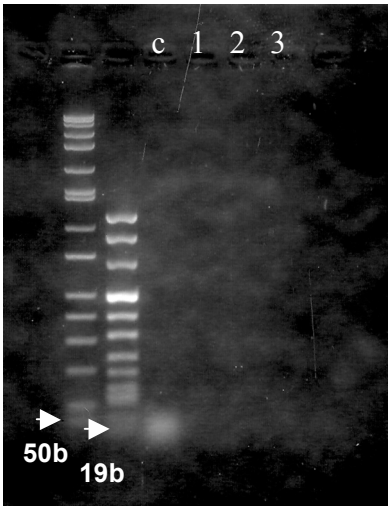


Figure 3.8: The *D. melanogaster* MSL3 chromodomain is likely to copurify with an oligonucleotide. **A.** After GST affinity purification, 1mL of the eluate was placed into a quartz cuvette and analyzed by UV spectroscopy, scanning absorption at wavelengths from 200 nm to 300 nm. **B.** 5% Agarose gel of the ethanol precipitated species. The first two lanes are molecular size markers followed by an untreated sample (control [lane c]), RNAse A treated (lanes 1 and 2) and benzonase treated samples (lane 3).

3.2.5 Site-directed Mutagenesis

In order to address the issue of the *D. melanogaster* MSL3 chromodomain forming oxidized cysteine oligomers, mutation of Cys66 was investigated (see section 2.3.16.2). Primers were designed for mutagenesis of the Cys residues to Ser, due to the conserved nature of the mutation.

Two peaks were observed during cation exchange chromatography of the *D. melanogaster* MSL3 chromodomain (2-91) C66S mutant (Figure 3.9). Both peaks are MSL3 (2-91) chromodomain, as determined by SDS-PAGE, indicating that the protein was still able to oxidize. It was concluded, based upon the C66S single mutation results, that keeping the protein in the presence of a reducing agent was more effective in preventing oxidation, than mutation of the Cys residues.

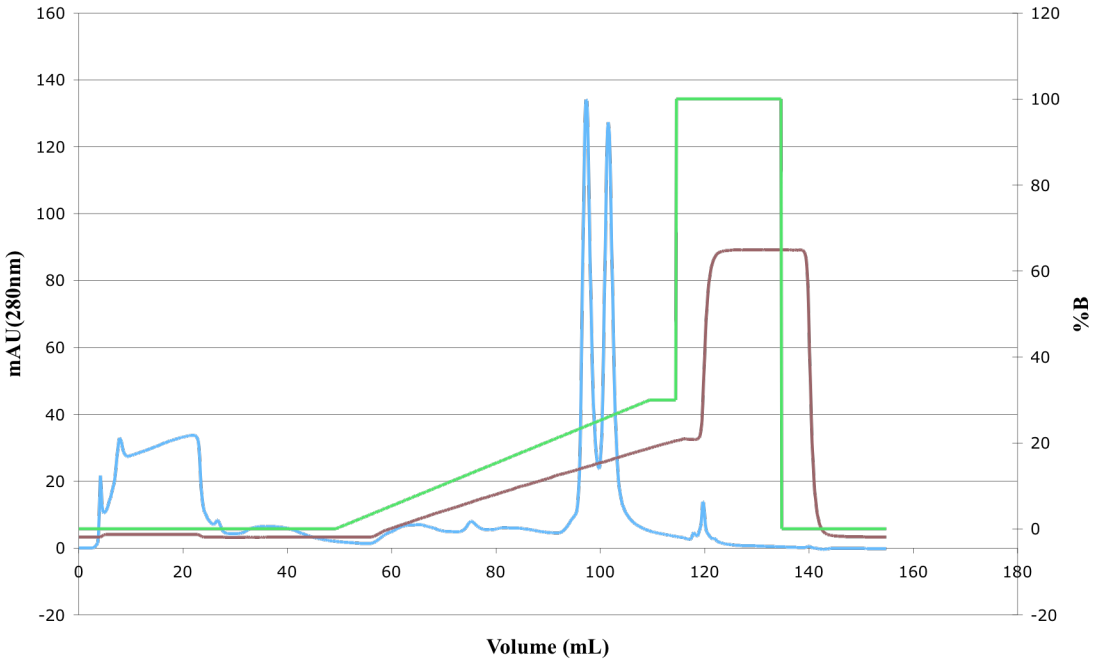
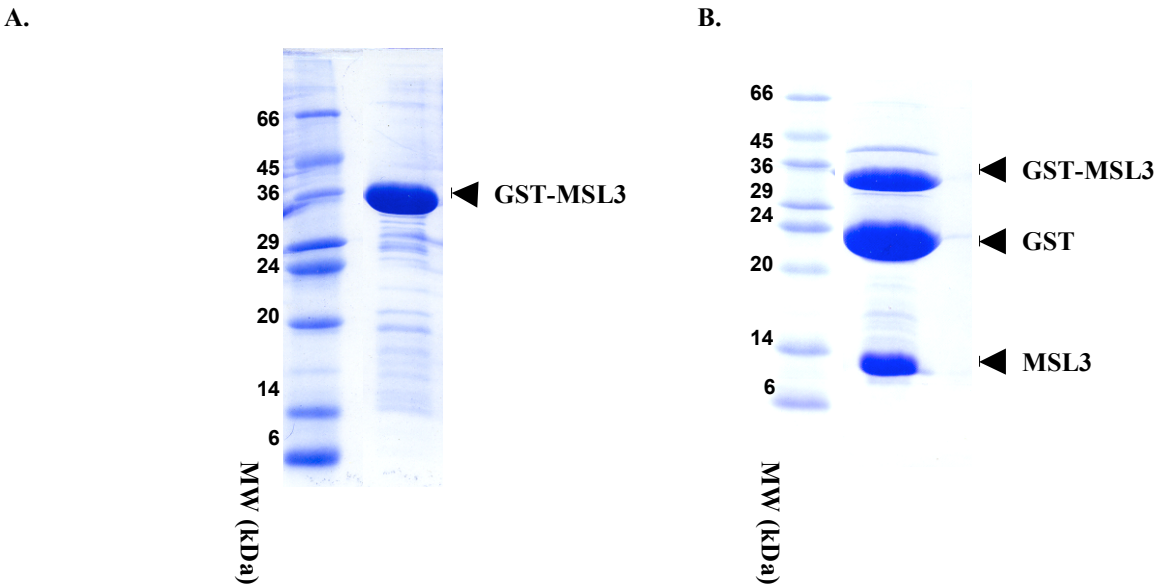


Figure 3.9: Cation exchange chromatogram of the *D. melanogaster* MSL3 chromodomain (2-91) mutant C66S. The mutant chromodomain was loaded onto a SourceS cation exchange column at 1 mL/min. The protein was bound and washed with buffer A prior to a gradient elution with buffer B. Absorbance (280 nm) is shown in blue (mAU is absorbing units x 1000), conductivity (indicating salt concentration) in brown (mS/cm), and concentration (percentage of buffer B) is shown in green (%B).

3.3 Cloning and Purification of the *H. sapiens* MSL3 Chromodomain (2-93)

The *H. sapiens* MSL3 (2-93) chromodomain construct was designed by Dr. S. Moore based upon sequence alignment with other known chromodomains, such as HP1 and Pc. Ewa Kerc (University of Saskatchewan) contributed to the cloning, ligation, transformation, and plasmid purification. The construct encodes a 92 amino acid domain with a theoretical pI and molecular weight of 9.72 and 10781.2 Da, respectively based upon the amino acid sequence. The *H. sapiens* MSL3 (2-93) chromodomain shares high sequence identity (46%) with its *D. melanogaster* orthologue (Figure 1.13). Purification of the *H. sapiens* MSL3 (2-93) chromodomain by GST affinity purification was effective in purifying the GST-MSL3 from the crude lysate. Only faint bands indicating remaining impurities were visible by SDS-PAGE. The GST-MSL3 ran at approximately 36 kDa, which was the expected size for the chimera (Figure 3.10 A). The chimera appeared stable as no visible degradation products were observed by SDS-PAGE. When digested with PreScission protease, three major fragments were observed corresponding to GST-MSL3 (36 kDa), GST (25 kDa), and MSL3 (11 kDa), indicating that cleavage had not gone to completion (Figure 3.10 B).



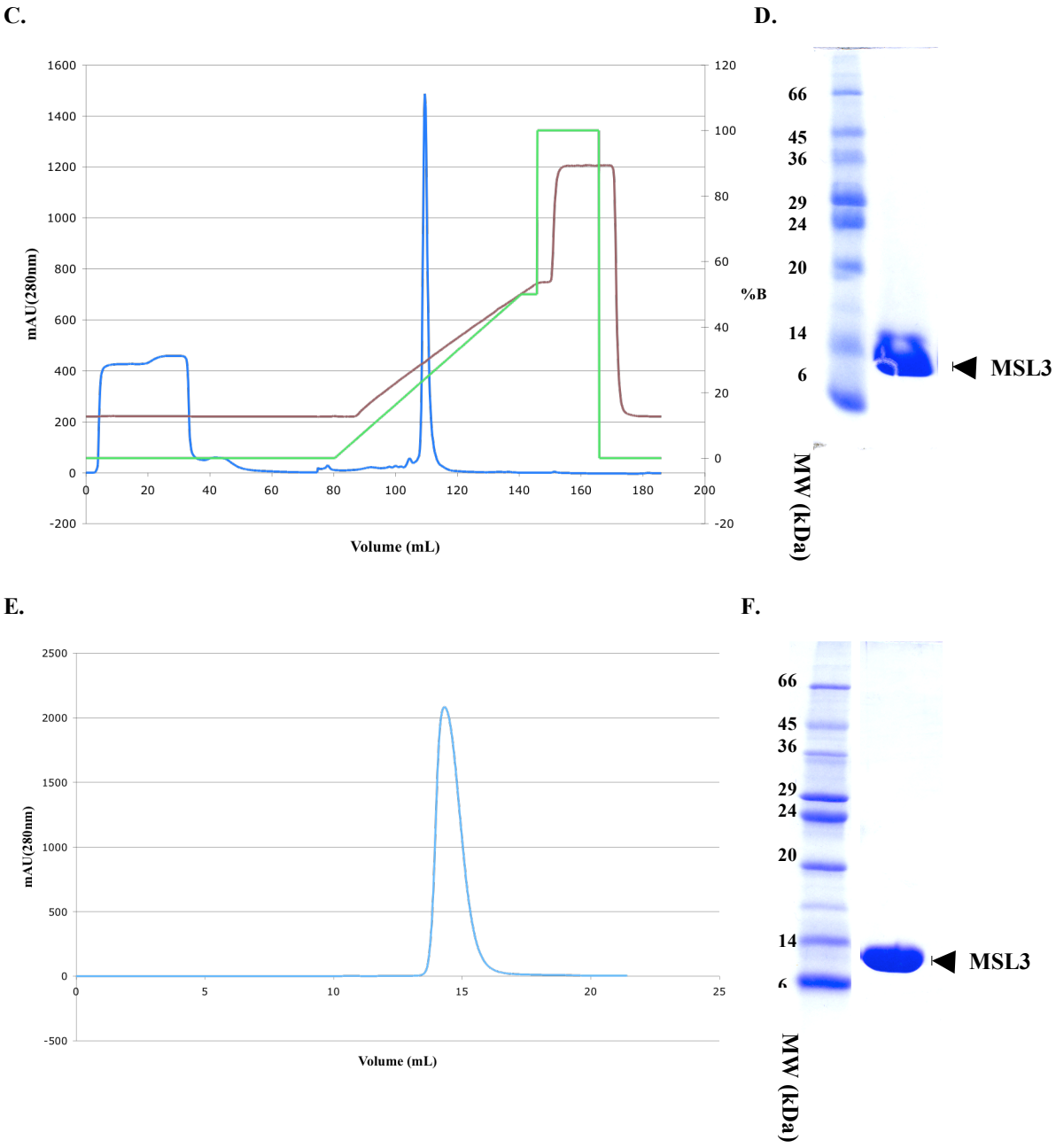


Figure 3.10: Purification of the *H. sapiens* MSL3 (2-93) chromodomain. **A.** SDS-PAGE (17% acrylamide:bis [37.5:1]) of the eluted GST-MSL3 after GST affinity purification. Molecular weight markers are shown. **B.** SDS-PAGE of the GST-MSL3 after cleavage with PreScission protease. **C.** Cation exchange chromatogram of the MSL3 chromodomain. The cleaved MSL3 chromodomain was loaded onto a SourceS cation exchange column at 1 mL/min. Absorbance (280 nm) is shown in blue (mAU is absorbing units x 1000), conductivity (indicating salt concentration) in brown (mS/cm), and concentration (percentage of buffer B) is shown in green (%B). **D.** SDS-PAGE of the peak eluted during cation exchange chromatography. **E.** Size exclusion chromatography of the MSL3 chromodomain. 500 µL was loaded onto a Superdex 75 column at a flow rate of 0.5 mL/min. The eluate was collected in 1 mL fractions. **F.** SDS-PAGE of the peak that eluted from the SEC75 column. Molecular weight markers are shown.

After GST affinity purification and PreScission protease digestion, the MSL3 (2-93) chromodomain was purified to homogeneity by cation exchange and size exclusion chromatography (SEC) (Figure 3.10 C-F). The *H. sapiens* MSL3 (2-93) chromodomain had a retention volume of 14.31 mL, which corresponds to an experimentally determined molecular weight of 11145 Da. The *H. sapiens* MSL3 (2-93) chromodomain was therefore homogeneous in solution as a monomer, based upon SEC and SDS-PAGE.

After size exclusion chromatography, the MSL3 (2-93) chromodomain was concentrated to approximately 20 mg/mL in a total volume of 200 μL, a total yield of approximately 4 mg of protein from 2 L of culture. The concentration was calculated using Beer's Law (Equation 3.1) and the solutions absorbance, measured at 280 nm.

3.3.1 Limited Proteolysis of the *H. sapiens* MSL3 (2-93) Chromodomain

In order to investigate the stability of the truncated *H. sapiens* MSL3 chromodomain (2-93), the purified protein was subjected to limited trypsin proteolysis (Figure 3.11). At a molar ratio of 1:1000 trypsin to MSL3 (2-93), proteolytic degradation was observed after 5 min at room temperature. A stable fragment was observed after an hour (Figure 3.11).

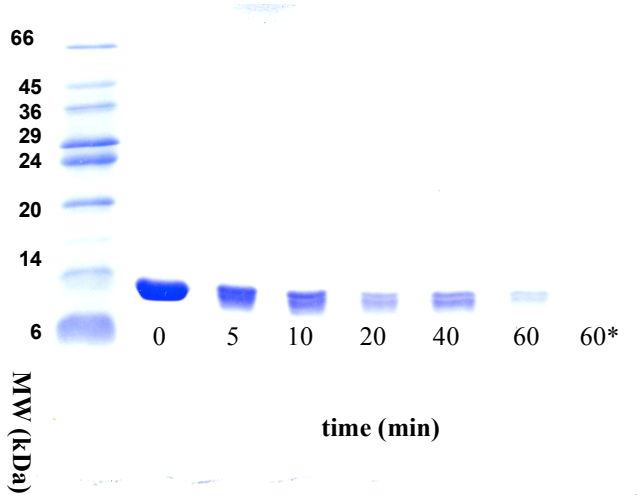


Figure 3.11: Limited proteolysis of the *H. sapiens* MSL3 chromodomain (2-93). hsMSL3 2-93 was incubated with trypsin at a 1:1000 molar ratio for 5, 10, 20, 40 and 60 minutes. A sample was also incubated at a molar ratio of 1:100 for 60 mins (60*).

Proteolytic degradation was complete, at a molar ratio of 1:100 after 1 hr. The conditions of the limited proteolysis experiment were identical to the proteolytic digestion of the *D. melanogaster* MSL3 (2-91) chromodomain. Comparison of the two indicated that the *H. sapiens* MSL3 (2-93) chromodomain was not likely to be as stable to trypsin digestion as its *D. melanogaster* counterpart.

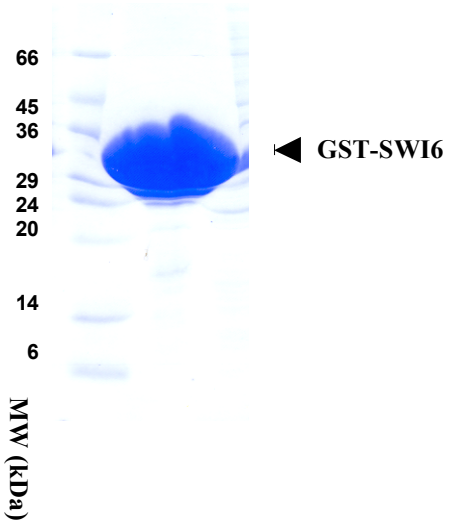
3.4 Purification and Crystallization Screening of the *S. pombe* SWI6 Chromodomain (77-135)

The *S. pombe* SWI6 (77-135) chromodomain was also cloned, expressed and purified, for use as a control in future binding studies. The *S. pombe* SWI6 (77-135) chromodomain is a close homologue of the *D. melanogaster* HP1 chromodomain, which is known to bind $H3K9Me_3$ histone N-terminal tails (Lachner *et al.*, 2001).

The *S. pombe* SWI6 (77-135) chromodomain construct encodes a 59 amino acid domain with a theoretical pI and molecular weight (MW) of 4.58 and 6959.5 Da, respectively, based upon the amino acid sequence. Ewa Kerc (University of Saskatchewan) contributed to the cloning, ligation, transformation, and plasmid purification.

Based upon SDS-PAGE, the GST affinity purification was effective in purifying the GST-SWI6 from the crude lysate, as no major other contaminants were observed. The GST-SWI6 ran at a molecular weight of approximately 32 kDa, as expected (Figure 3.12 A). The SWI6 (77-135) chromodomain was purified to homogeneity by anion exchange and SEC (Figure 3.13 B-E). Anion exchange was chosen due to the low pI of the protein (4.58).

A.



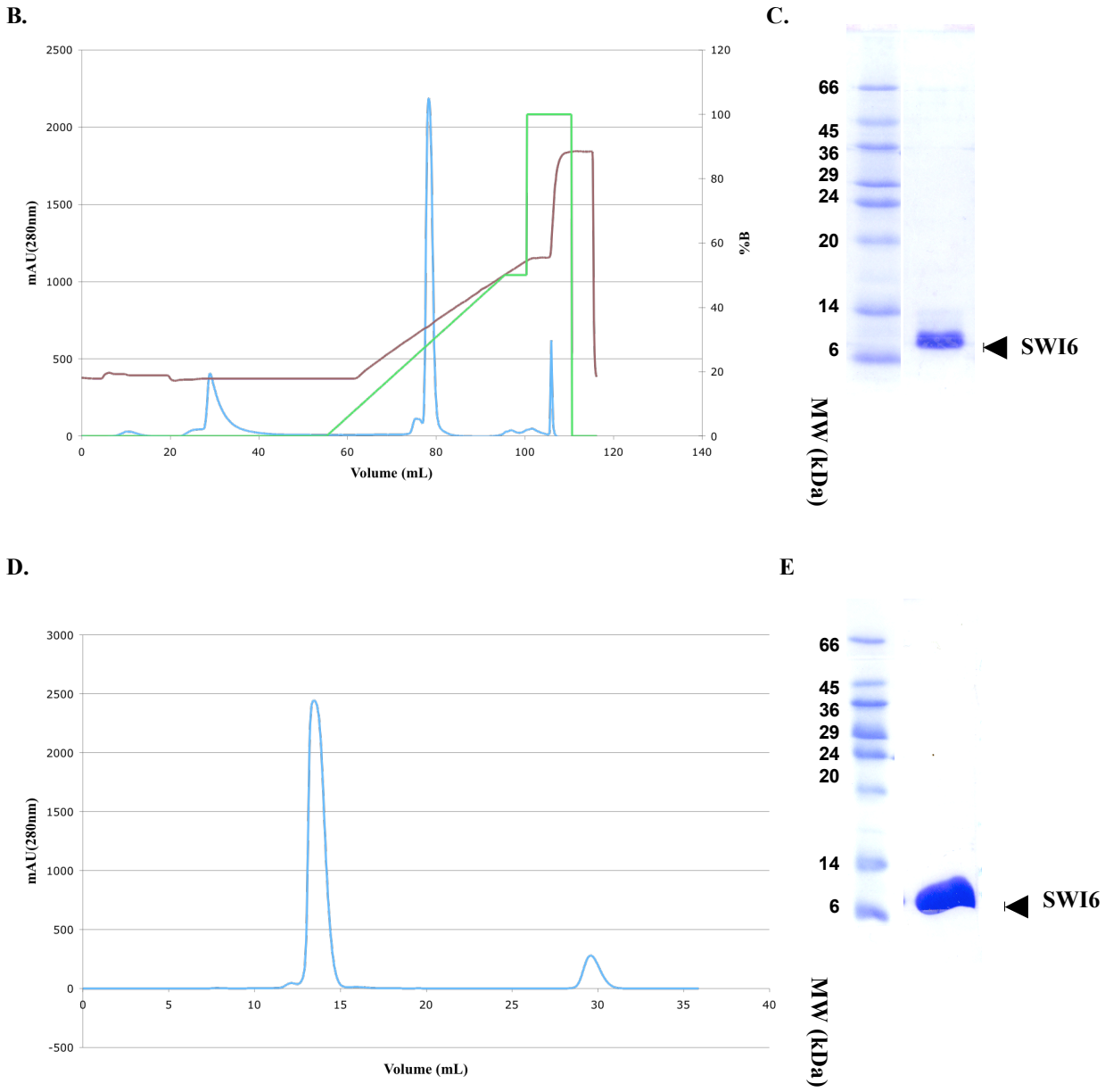


Figure 3.12: Purification of the *S. pombe* SWI6 (77-135) chromodomain. **A.** SDS-PAGE (17% acrylamide:bis [37.5:1]) of the eluted GST-SWI6 after GST affinity purification. **B.** Anion exchange chromatogram of the SWI6 chromodomain. The cleaved SWI6 chromodomain was loaded onto a SourceQ anion exchange column at 1 mL/min. The protein was bound and washed with buffer A (see section 2.3.13.1) prior to a gradient elution with buffer B. Absorbance (280 nm) is shown in blue (mAU is absorbing units x 1000), conductivity (indicating salt concentration) in brown (mS/cm), and concentration (percentage of buffer B) is shown in green (%B). **C.** SDS-PAGE of the peak eluted during anion exchange chromatography. **D.** Size exclusion chromatography of the SWI6 chromodomain. 500 µL was loaded onto a Superdex 75 column at a flow rate of 0.5 mL/min. The eluate was collected in 1 mL fractions. **E.** SDS-PAGE of the peak that eluted from the SEC75 column.

The SWI6 (77-135) chromodomain has a theoretical molecular weight of 6959 Da and an apparent molecular weight of 7 kDa, as determined by SDS-PAGE. The SWI6 (77-135) chromodomain eluted from the Superdex 75 column at a retention volume of 13.47 mL, which corresponded to a molecular weight of 15845 Da, calculated using the standard curve. This may indicate that the SWI6 (77-135) chromodomain is a dimer in solution. The purified SWI6 (77-135) chromodomain was concentrated to approximately 20 mg/mL, in a total volume of 400 μL, a total yield of approximately 8 mg of protein from 2 L of culture.

3.4.1 Vapour Diffusion Crystallization Screens

After ion exchange and SEC chromatography, the purified *S. pombe* SWI6 (77-135) chromodomain was concentrated to approximately 20 mg/mL using Millipore Amicon Ultra 5k Dalton cut-off centrifugal filter devices and equilibrated into a buffer containing 20 mM Bis-Tris-propane, pH 7, 50 mM NaCl, 5 mM TCEP. Crystallization conditions were initially investigated using the Emerald Biosciences Wizard I and II and Hampton Research Crystal Screen I and II commercial sparse matrix screens. For each condition, two drops were set up, one with just the SWI6 chromodomain and one with the chromodomain and the H3K9Me$_3$ peptide, at a 20% molar excess. No conditions were identified that yielded observable crystals.

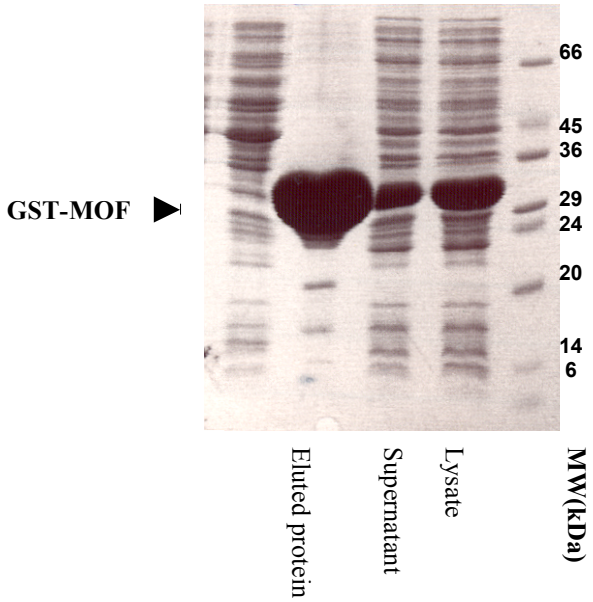
3.5 *D. melanogaster* MOF Chromodomain (371-440)

3.5.1 Cloning and Purification of the *D. melanogaster* MOF Chromodomain (371-440)

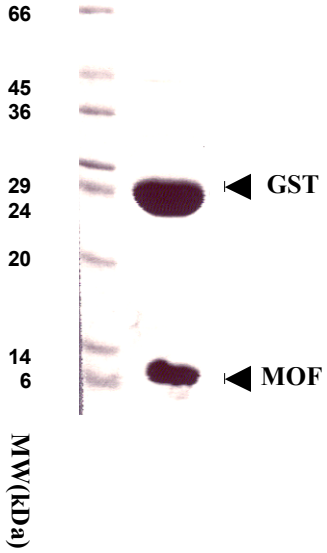
The original *D. melanogaster* MOF chromodomain (371-440) construct was designed by Dr. S. Moore and cloned into the pGEX-6P3 expression vector by J. Guenther (University of Saskatchewan). The theoretical pI and molecular weight of the 70 amino acid MOF (371-440) chromodomain is 5.84 and 8098.8 Da, respectively.

After GST affinity chromatography, the eluted GST *D. melanogaster* MOF (371-440) chromodomain was analyzed by SDS-PAGE; a large single band was observed at approximately 33 kDa. The GST affinity purification was effective in purifying the GST-MOF from the crude lysate as no other major bands were observed by SDS-PAGE (Figure 3.13 A). When digested with PreScission protease, two fragments were observed on the SDS-PAGE gel,

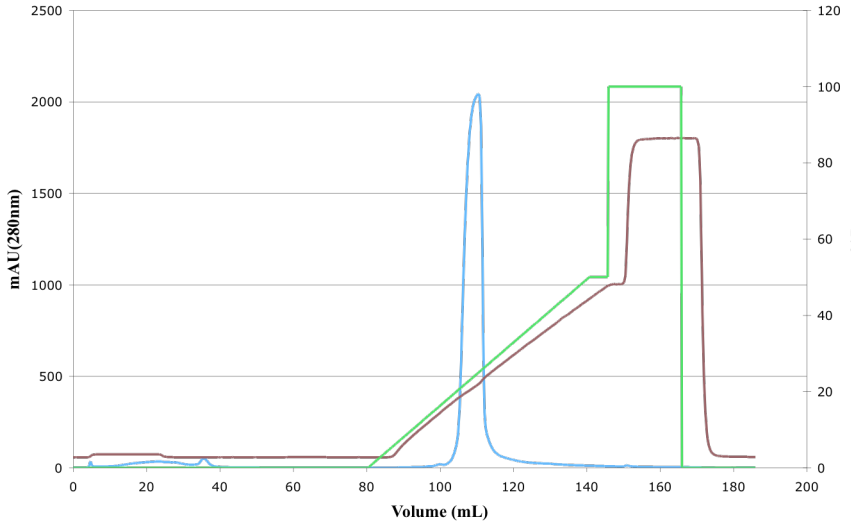
A.


GST-MOF
66
45
36
29
24
20
14
6
Eluted protein
Supernatant
Lysate
MW(kDa)

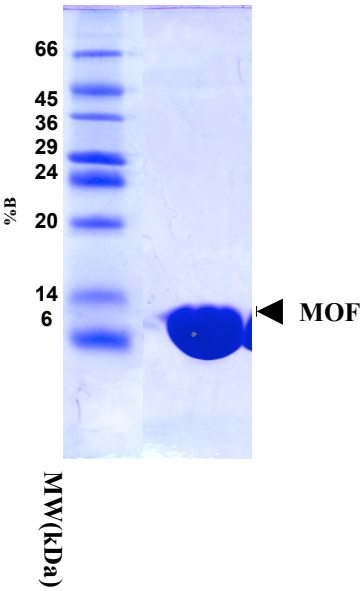
B.


66
45
36
29
24
20
14
6
GST
MOF
MW(kDa)

C.


mAU(280nm)
%B
Volume (mL)
2500
2000
1500
1000
500
0
120
100
80
60
40
20
0
0
20
40
60
80
100
120
140
160
180
200

D.


66
45
36
29
24
20
14
6
MOF
MW(kDa)

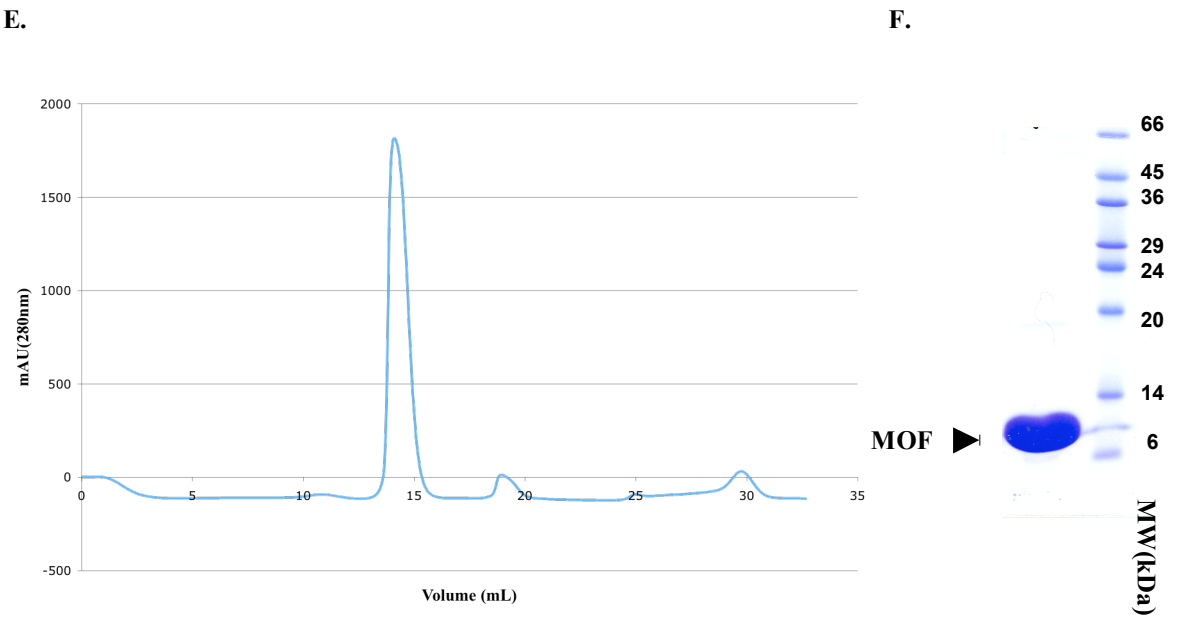


Figure 3.13: Purification of the *D. melanogaster* MOF chromodomain (371-440). **A.** SDS-PAGE (17% acrylamide:bis [37.5:1]) of the eluted GST-MOF after GST affinity purification. Molecular weight markers are shown. **B.** SDS-PAGE of the GST-MOF after cleavage with PreScission protease. **C.** Cation exchange chromatogram of the MOF chromodomain. The cleaved MOF chromodomain was loaded onto a SourceS cation exchange column at 1 mL/min. The protein was bound and washed with buffer A (see section 2.3.13.1) prior to a gradient elution with buffer B. Absorbance (280 nm) is shown in blue (mAU is absorbing units x 1000), conductivity (indicating salt concentration) in brown (mS/cm), and concentration (percentage of buffer B) is shown in green (%B). **D.** SDS-PAGE of the MOF chromodomain after cation exchange chromatography. **E.** Size exclusion chromatography of the MOF chromodomain after cation exchange chromatography. 500 µL was loaded onto a Superdex 75 column at a flow rate of 0.5 mL/min. The eluate was collected in 1 mL fractions. **F.** SDS-PAGE of the peak that eluted from the SEC75 column. Molecular weight markers are shown.

as expected, with molecular weights consistent with the MOF (371-440) chromodomain (8 kDa) and GST (25 kDa) (Figure 3.14 B). After removal of the cleaved GST and further purification by cation exchange chromatography and SEC, the *D. melanogaster* MOF chromodomain (371-440) was visibly homogenous, based upon SDS-PAGE (Figure 3.13 F).

The theoretical molecular weight of the *D. melanogaster* MOF (371-440) chromodomain is 8098.8 Da. From the SEC75 chromatogram, the MOF (371-440) chromodomain had a retention volume of 14.11 mL, which corresponded to a molecular weight of 12119 Da. The *D. melanogaster* MOF (371-440) chromodomain was therefore likely to be homogenous in solution as a monomer.

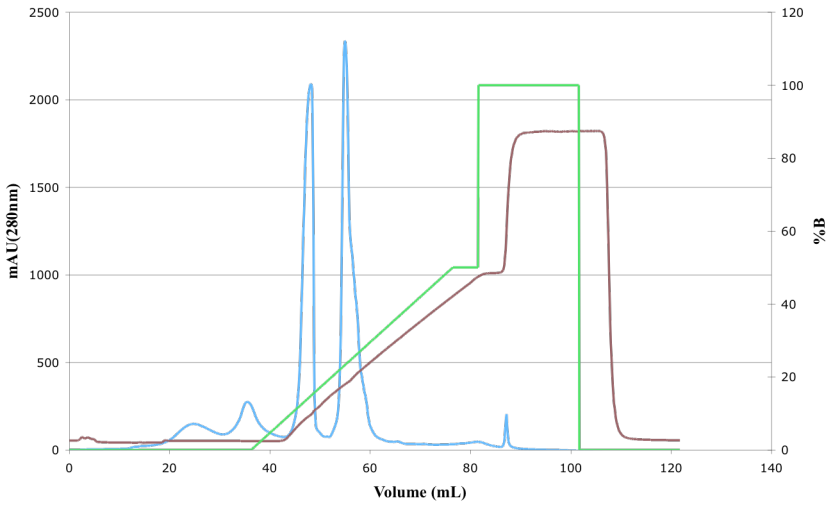
The *D. melanogaster* MOF precipitated irreversibly when dialysed from elution buffer (pH 8.0) into a sodium acetate buffer (pH 5.0) for cation exchange, through its isoelectric point (pI 5.84). This was only partially abrogated by dialyzing in two steps, as previously described (section 2.3.14). The theoretical pI of the *D, melanogaster* MOF chromodomain (371-440), based upon the amino acid sequence is 5.84 (70 amino acids, MW 8098.8 Da). The *D. melanogaster* MOF chromodomain (371-440) was purified by anion exchange in a buffer containing 50 mM Tris·HCl, 5 mM DTT, pH 8.0 (buffer A) (Figure 3.14 A). No precipitation was observed during dialysis. The two peaks observed in the chromatogram are *D. melanogaster* MOF (371-440) chromodomain (left) and GST (right). The remaining GST is likely due to insufficient dialysis to remove the majority of the glutathione. Compared to the SEC75 chromatogram of the MOF (371-440) chromodomain purified by cation exchange, the SEC75 profile is unchanged, as expected (Figure 3.14 B).

3.5.2 Vapour Diffusion Crystallization Screens

After ion exchange and SEC chromatography, the purified *D. melanogaster* MOF chromodomain (371-440) was concentrated to approximately 20 mg/mL using Millipore Amicon Ultra 5k Dalton cut-off centrifugal filter devices and equilibrated into a buffer containing 20 mM Bis-Tris-propane, pH 7, 50 mM NaCl, 5mM TCEP. Crystallization conditions were initially investigated using the Emerald Biosciences Wizard I and II and Hampton Research Crystal Screen I and II commercial sparse matrix screens. Small crystals were observed in condition 16 (1.0 M sodium citrate, 100 mM CHES, pH 9.5) of the Wizard II screen. The crystals had a hexagonal column morphology, indicating a possible hexagonal or trigonal Bravais lattice (Figure 3.15). Sodium citrate concentration and pH were screened to optimize crystal size. Crystals of sufficient size (approximately 0.5 mm by 0.1 mm) for diffraction on the in-house Proteum Copper (Cu) rotating anode diffractometer were grown from 1 M sodium citrate, 100 mM CHES, pH 9.5, 15% glycerol (Figure 3.16). The inclusion of

glycerol meant that the crystals could be directly frozen in the mother liquor (the solution from which the crystals grew) without requiring transfer into a cryo-protectant.

A.



B.

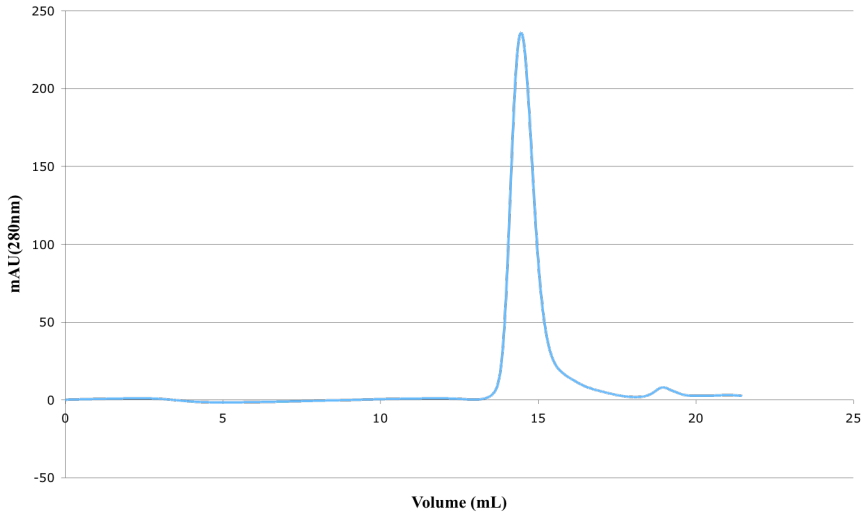


Figure 3.14: Anion exchange chromatography of the *D. melanogaster* MOF (371-440) chromodomain. **A.** The cleaved MOF chromodomain was loaded onto a SourceQ anion exchange column at 1 mL/min. The protein was bound and washed with buffer A (see section 2.3.13.1) prior to a gradient elution with buffer B. Absorbance (280 nm) is shown in blue (mAU is absorbing units x 1000), conductivity (indicating salt concentration) in brown (mS/cm), and concentration (percentage of buffer B) is shown in green (%B). **B.** Size exclusion chromatography of the MOF chromodomain after pooling and concentration of the fractions from anion exchange. 500 µL was loaded onto a Superdex 75 column at a flow rate of 0.5 mL/min. The eluate was collected in 1 mL fractions.

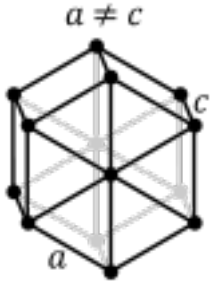
3.5.3 Data Collection and Data Processing of Native *D. melanogaster* MOF Chromodomain (371-440) Crystals

Initial diffraction images and the first native data set were collected at the Saskatchewan Structural Sciences Centre using a DX8 Proteum Copper (Cu) rotating anode diffractometer with a Kappa goniostat and Proteum 4K charge-coupled device (CCD) detector. Crystals were diffracted under a stream of nitrogen at 114 K using a Kryoflex low temperature apparatus. The crystals diffracted well (Figure 3.16). Spots were discernible out to 1.9 Å resolution. Each image was collected over a 0.5° oscillation around omega over a 1 min period. A total of 600 images, 300° of data were collected. The diffraction images were processed using SaintPlus and ProScale (Bruker AXS Inc). The cell parameters and the statistics of the processed data are summarized in Tables 3.2 and 3.3, respectively.

The unit cell was indexed and the cell parameters calculated. Based upon the cell volume (V_{cell}), using the Matthews coefficient equation, there are likely four or five molecules in the asymmetric unit of the primitive hexagonal lattice (Equation 3.3) (Matthews, 1968) (Figure 3.17).

A.

Volume = $a^2c(\sin[60^o])$



B.

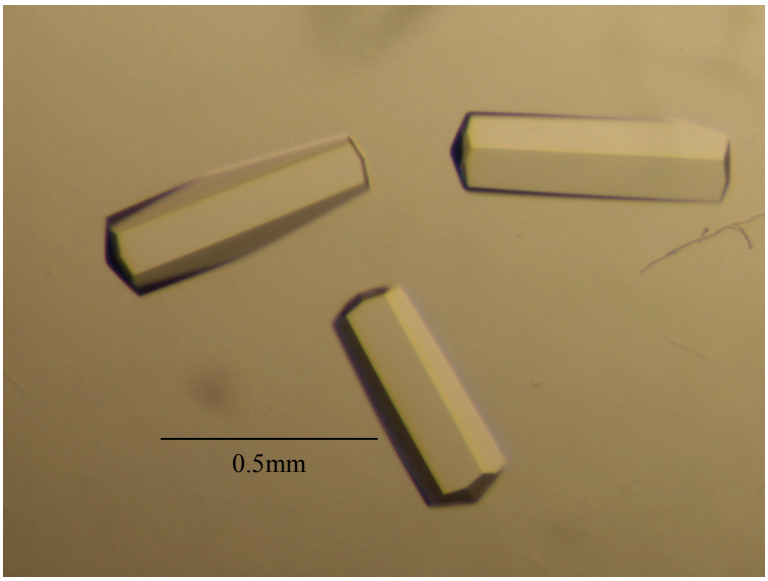


Figure 3.15: Crystals of *D. melanogaster* MOF (371-440) chromodomain. **A.** A schematic diagram depicting the hexagonal Bravais lattice and the formula for calculating the unit cell volume. **B.** Crystals of *D. melanogaster* MOF (371-440) chromodomain, grown from 1 M sodium citrate, 100 mM CHES, pH 9.2, 15% glycerol.

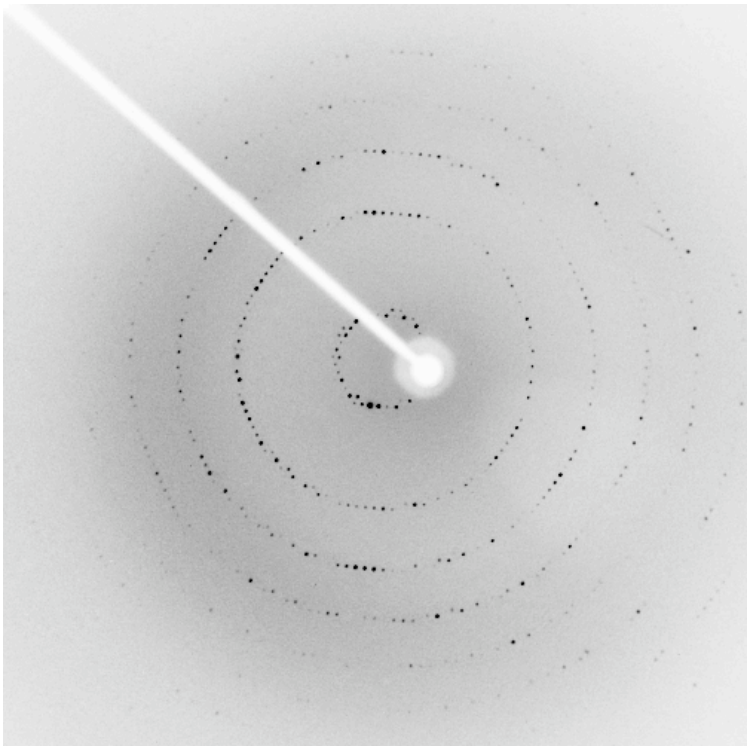


Figure 3.16: Bruker Proteum CCD detector diffraction image of *D. melanogaster* MOF chromodomain crystals taken over a 0.5° oscillation for 1 min.

Table 3.2: Summary of the native *D. melanogaster* MOF chromodomain crystals unit cell dimensions. The Bravais lattice is hexagonal and is either $P6_1$ or $P6_5$.

Unit cell	
a	113.9 Å
b	113.9 Å
c	50.4 Å
α	90^o
β	90^o
γ	120^o
Hexagonal	$P6_1$ or $P6_5$
Volume (V_{cell})	566250 $Å^3$
Mosaicity	0.45^o

Table 3.3: Summary of data processing for the native *D. melanogaster* MOF chromodomain (371-440) crystal data set. The resolution of each shell (Res. Shell), the number of observed reflections (nRef Obs), the number of expected reflections (nRef Exp), the completeness (Comp. %), redundancy (Redund), the difference between multiple symmetry related reflections (Rsym), the R factor applied to a small subset of reflections that are not used in the refinement (Rfree), the measure of the reflections intensity compared to the background (I/sigI), and the statistical measure of the fit between the theoretical frequency distribution and the observed frequency distribution of the observed data (Chi^2) are shown.

	Res. Shell	nRef Obs	nRef Exp	Compl %	Redund	Rsym-	shell	Rfree	nRfree	I/sigI	shell	Chi^2
	37.88											
to	4.82	1898	1910	99.37	10.79	0.0343	0.0343	0.0381	86	20.51	20.51	1.11
to	3.81	3809	3821	99.69	10.93	0.0365	0.038	0.0397	187	19.84	19.2	1.14
to	3.32	5720	5738	99.69	10.97	0.0407	0.0517	0.0428	280	17.8	13.75	1.11
to	3.01	7632	7651	99.75	10.83	0.0459	0.0767	0.0478	376	15.8	9.54	1.08
to	2.79	9529	9551	99.77	10.57	0.0513	0.12	0.0535	464	14.06	6.08	1.04
to	2.62	11451	11474	99.8	10.2	0.0564	0.174	0.0581	557	12.7	4.18	1
to	2.49	13351	13380	99.78	9.54	0.0598	0.2104	0.0608	637	11.94	3.5	0.98
to	2.38	15258	15287	99.81	8.87	0.0623	0.255	0.0635	732	11.41	2.9	0.95
to	2.29	17181	17215	99.8	8.24	0.0641	0.2562	0.0661	838	11.03	2.71	0.93
to	2.2	19096	19384	98.51	7.67	0.0655	0.2737	0.0674	912	10.74	2.39	0.91
to	2.1	21017	22173	94.79	7.14	0.0664	0.332	0.0682	966	10.52	1.74	0.9

$$\mathbf{Vm} = \frac{\mathbf{V_{cell}}}{\mathbf{M_w(Da)*Z*N}}$$

Equation 3.3: Matthew's coefficient (V_M) equation. Z is the number of asymmetric units in the unit cell, N is the number of molecules in the asymmetric unit and M_w(Da) is the molecular weight of the molecule.

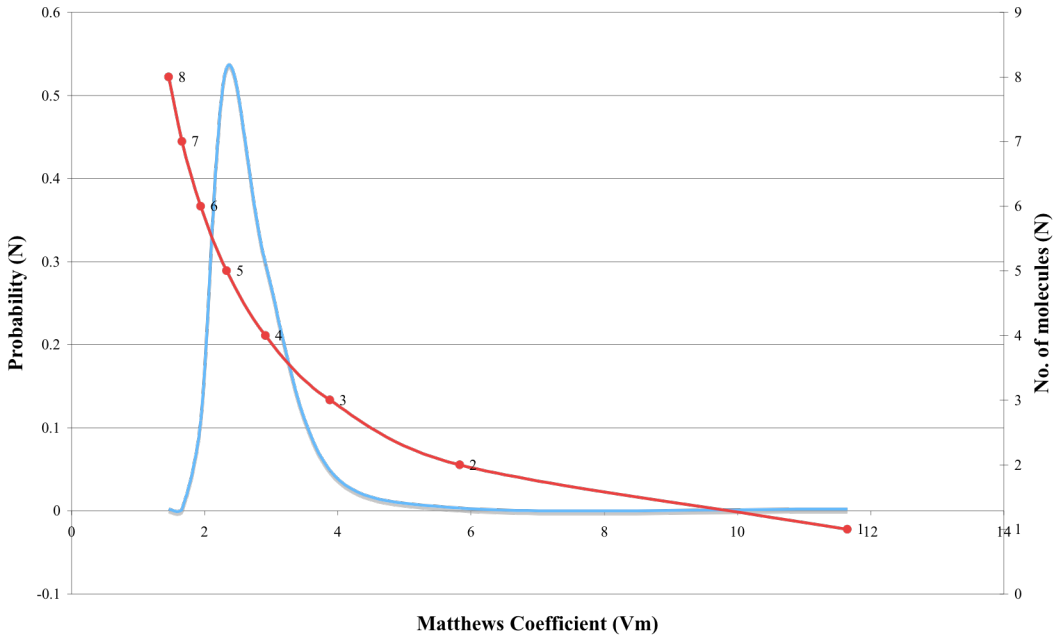


Figure 3.17: Matthews's coefficient (V_M) calculation for the *D. melanogaster* MOF chromodomain (371-440) native data set. The probability (blue) of the number of molecules in the asymmetric unit (red [X]) is plotted against the Matthews coefficient (Vm). Z is the number of asymmetric units in the unit cell (Matthews, 1968).

3.5.4 Molecular Replacement.

Molecular replacement was attempted to solve the *D. melanogaster* MOF chromodomain (371-440) crystal structure using the native data set, the *D. melanogaster* MOF chromodomain NMR structure (Nielsen *et al.*, 2005) and the PHASER molecular replacement software (McCoy *et al.*, 2007), as outlined in section 2.3.22.1. The solution structure protein data base file (2BUD) is an ensemble of 25 models. The whole or partial ensemble was used in PHASER to calculate structure factors with the full side-chain (Full) or the side-chain truncated to an Ala residue (Ala) with the exception of Gly. Analysis of the ensemble reveals large root-mean-squared deviation of residues at the N- and C-terminus between models. Therefore in all cases each model used in the respective ensemble was truncated to residues 378 to 443. An ensemble was also prepared with the 402-410 loop excised due to large root-mean-squared deviation between models (Ala-loop). The PHASER results are summarized in Table 3.4.

Table 3.4: Summary of the molecular replacement results using PHASER (McCoy *et al*., 2007) and the *D. melanogaster* MOF chromodomain NMR structure (Nielsen *et al*., 2005). The search model ensemble either had the residue side chain (full) or was truncated to an Ala. The ensemble number is the number of models used to calculate the model structure factors. The resolution is the data used in the calculation and the search is the number of molecules searched for in the asymmetric unit. Log-likelihood gain (LLG) statistics and the Z-score (signal to noise) are shown for the initial and final fast rotation function (FRF) and fast translation (FTF) solutions. Eulerian angles were not consistent for each of the solution.

				Initial Sol.			**Final Sol.**	
	Ensemble	**RESOLution**	**Search**	**FRF** LLG (Z)	**FTF** LLG (Z)		**FRF** LLG (Z)	**FTF** LLG (Z)
Full	25	4-8	2	4.79 (3.17)	12.64 (4.70)	P65	18.59 (3.20)	29.90 (4.84)
Full	1	4-8	2	4.05 (3.55)	9.85 (4.40)	P61	15.55 (3.44)	27.45 (5.28)
Full	23	6-15	2	4.61 (3.87)	10.01 (3.36)	P65	15.04 (3.52)	27.27 (4.97)
Full	23	6-10	4	3.67 (3.76)	8.82 (3.68)	P65	46.42 (3.34)	56.45 (4.70)
Full	23	6-37	1	4.77 (3.86)	10.12 (3.45)	P61		
Full	23	6-37	2	15.94 (3.51)	23.84 (4.54)	P61	33.19 (3.04)	44.17 (4.54)
Ala-loop	24	4-8	2	4.46 (3.82)	9.98 (3.79)	P65	16.18 (4.03	26.86 (4.77)
Ala-loop	24	4-8	4	4.46 (3.82)	9.98 (3.79)	P65	54.34 (3.37)	66.02 (4.68)
Ala-loop	24	6-37	3	3.19 (3.16)	8.41 (3.83)	P61	32.31 (3.17)	43.21 (4.23)
Ala-loop	24	4-37	3	6.41 (3.87)	14.51 (4.44)	P61	48.55 (4.12)	59.51 (4.83)
Ala-loop	24	4-10	3	5.89 (3.85)	13.81 (4.45)	P65	42.50 (3.68)	52.94 (4.55)
Ala-loop	24	4-10	1	3.14 (3.88)	8.38 (4.66)	P65		
Ala	25	3-15	4	9.60 (3.89)	19.18 (4.59)	P65	55.51 (3.95)	22.17 (5.15)
Full	6	4-8	4	8.914 (3.48)	19.74 (4.88)	P61	75.11 (3.83)	57.52 (5.06)
Full	6	4-8	2	6.00 (4.44)	12.73 (4.36)	P61	19.22 (4.07)	33.49 (5.61)

The log-likelihood gain (LLG) statistics are very poor with no clear solution; the best solution had a LLG of 9.98 after the initial rotation and translation functions, increasing to only 66.02 after the fourth molecule was fit. In comparison, the crystal structure of the Thrombin-activable Fibrinolysis Inhibitor (TAFI), which has three molecules in the asymmetric unit had a LLG of 367 for the initial solution, and a LLG of 2047 for the third (Anand *et al*., 2008).

3.5.5 Site-Directed Mutagenesis, Purification and Crystallization of Se-Met L423M Mutant *D. melanogaster* MOF Chromodomain (371-440)

Failure of molecular replacement to find a solution using PHASER (McCoy *et al*., 2007) and the solution structure of the *D. melanogaster* MOF chromodomain, solved by NMR (Nielsen *et al*., 2005)(PDB 2bud) meant that experimental techniques to measure the phases needed to be investigated.

The *D. melanogaster* MOF chromodomain (371-440) construct does not contain any methionine residues. In order to incorporate seleno-methionine (Se-Met) into the MOF (371-440) chromodomain to allow determination of phases by multi-wavelength anomalous dispersion (MAD), a methionine residue had to be introduced, via site-directed mutagenesis. A total of nine methionine mutants were cloned, expressed, and purified. Attempts to crystallize the purified I373M, L419M, and L423M *D. melanogaster* MOF chromodomain (371-440) mutants were made by screening sodium citrate concentration and pH around the conditions under which the wild-type protein crystallized. Results for each of the mutants is summarized in Table 3.5.

The Se-Met MOF L423M mutant was purified under conditions similar to purification of the wild-type protein, using anion exchange chromatography to minimize the amount of protein lost to irreversible precipitation during dialysis. Overlay of the Se-Met MOF L423M mutant SEC75 profile with that of the L423M mutant and the wild-type protein revealed a small variation in the elution volume of the three proteins (Figure 3.18). A small variation in V_e is often observed, therefore the profiles of the three proteins are essentially identical. The Se-Met L423M MOF (371-440) crystals exhibited a hexagonal column morphology identical to the wild type crystals (Figure 3.19).

Table 3.5: Summary of each of the *D. melanogaster* MOF chromodomain (371-440) Met mutants. Only the L423M mutant was able to be crystallized under conditions close to the wild-type protein. (-) indicates that the protein was not crystallized or that a Se-Met preparation was not carried out.

Mutant	Expressed	Purified	Crystallized	Se-Met Prep.
I373M	yes	yes	-	-
I385M	yes	yes	-	-
V397M	yes	yes	-	-
V414M	yes	yes	-	-
L419M	yes	yes	-	-
L423M	yes	yes	yes	yes
V427M	yes	yes	-	-
I432M	yes	yes	-	-
L434M	yes	yes	-	-

3.5.6 Data Collection and Data Processing of the Se-Met *D. melanogaster* MOF Chromodomain (371-440) L423M Crystals

Initial diffraction images were collected at the Saskatchewan Structural Sciences Centre using a DX8 Proteum rotating Copper (Cu) anode diffractometer with a Kappa goniostat and Proteum 4K charge-coupled device (CCD) detector. Crystals were diffracted under a stream of nitrogen at 114 K using a Kryoflex low temperature apparatus. The crystals diffracted poorly (Figure 3.20). Diffraction was only discernible to approximately 3.3 Å. A total of 50 frames with a 0.5° oscillation around omega, over a 1 min period were collected.

Fresh Se-Met *D. melanogaster* L423M MOF (371-440) chromodomain mutant protein was prepared and crystallized for diffraction experiments at the Canadian Light Source (CLS) on the Canadian Macromolecular Facility 1 (CMCF1) 08ID-1 beamline using a MAR225 Detector. A selenium fluorescence scan was carried out to ascertain if Se-Met had been incorporated (Figure 3.21). Data was collected at the peak (12.6558 keV) and inflection (12.6553) points of the selenium absorption edge.

A.

Protein	Ve	Kav	MW
Wild type	14.46	0.4097	**10447.2**
L423M Met	13.97	0.3793	**12852.9**
L423M Se-Met	14.19	0.3929	**11721.9**

B.

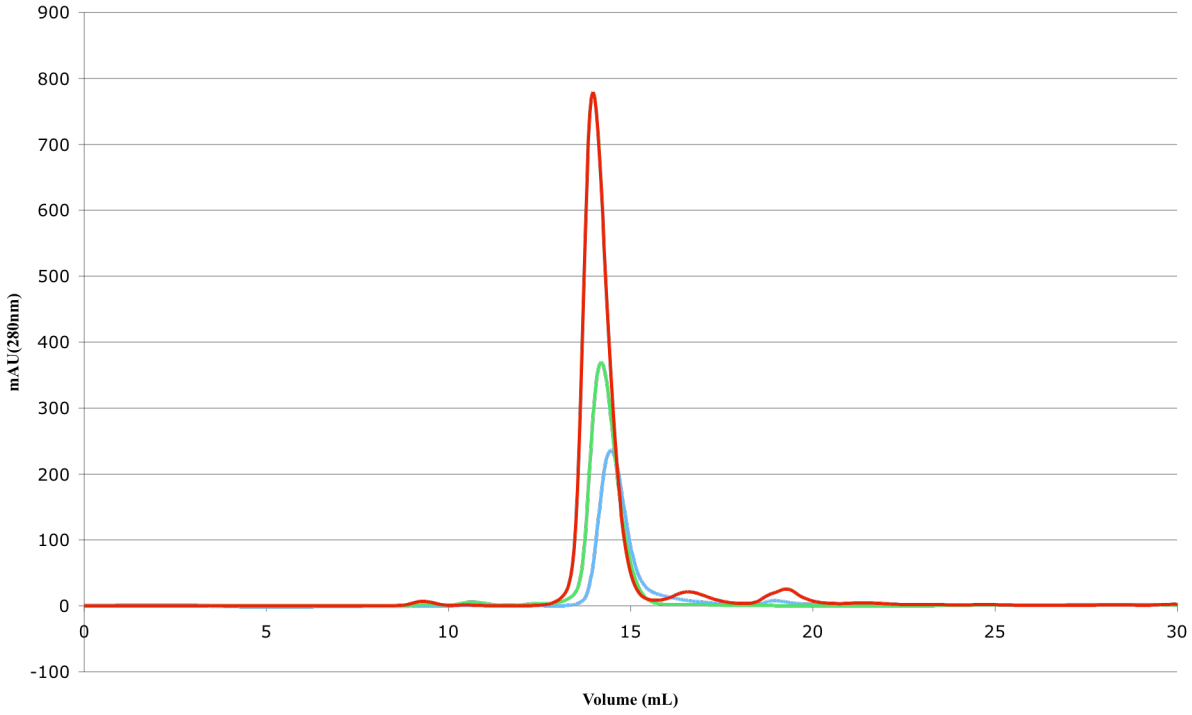


Figure 3.18: Comparison of the wild type (blue), L423M Met mutant (red), and L423M Se-Met mutant (green) *D. melanogaster* MOF (371-440) chromodomain SEC75 chromatograms. The respective retention volumes (Ve) and experimental molecular weights (MW) are shown (**A)**, along with the SEC75 chromatograms of the three proteins (**B)**.

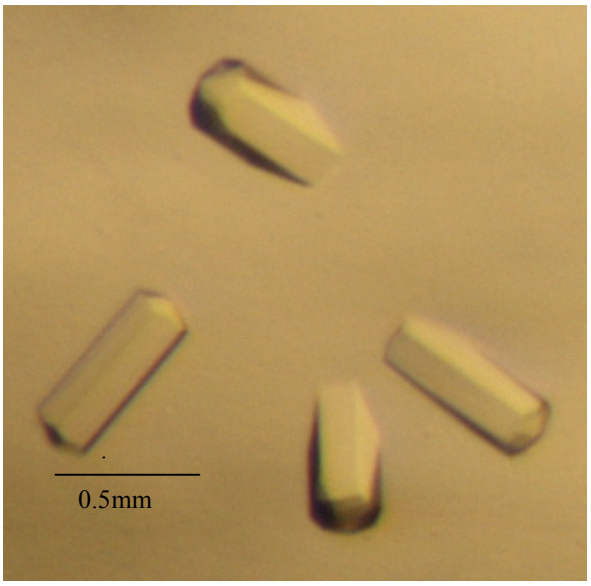


Figure 3.19: Crystals of the Se-Met derivative of the *D. melanogaster* MOF chromodomain (371-440) L423M mutant.

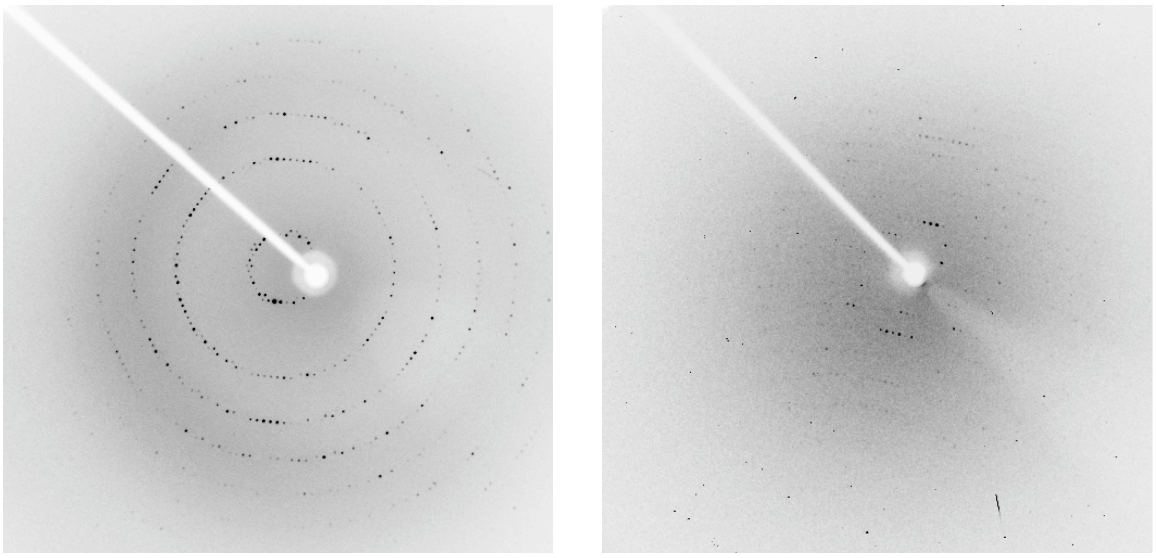


Figure 3.20: Comparison of the diffraction from wild type (left) and Se-Met derivative *D. melanogaster* MOF chromodomain (371-440) (right) crystals, taken over a 0.5° oscillation for 1 min. The images were collected using DX8 Proteum rotating Copper (Cu) anode diffractometer with a Kappa goniostat and Proteum 4K charge-coupled device (CCD) detector.

Diffraction of the crystals on the CMCF1 08ID-1 beamline was of much better quality than the diffraction from the DX8 Proteum rotating Copper (Cu) anode diffractometer. The diffraction images were processed using the HKL2000 (Otwinowski and Minor, 1997). Indexing of initial frames indicated that contrary to the earlier results, the Se-Met unit cell was in fact isomorphous to the wild type unit cell. The earlier result indicating that it was not isomorphous is likely due to a failure to index correctly due to the poor nature of the diffraction.

A large number of crystals were mounted and diffracted. A large variation in the quality of diffraction was observed. A total of five Se-Met MAD data sets were collected. The data was systematically processed to optimize their relative statistics and the individual unmerged data sets compared. Processing results are summarized in Table 3.6.

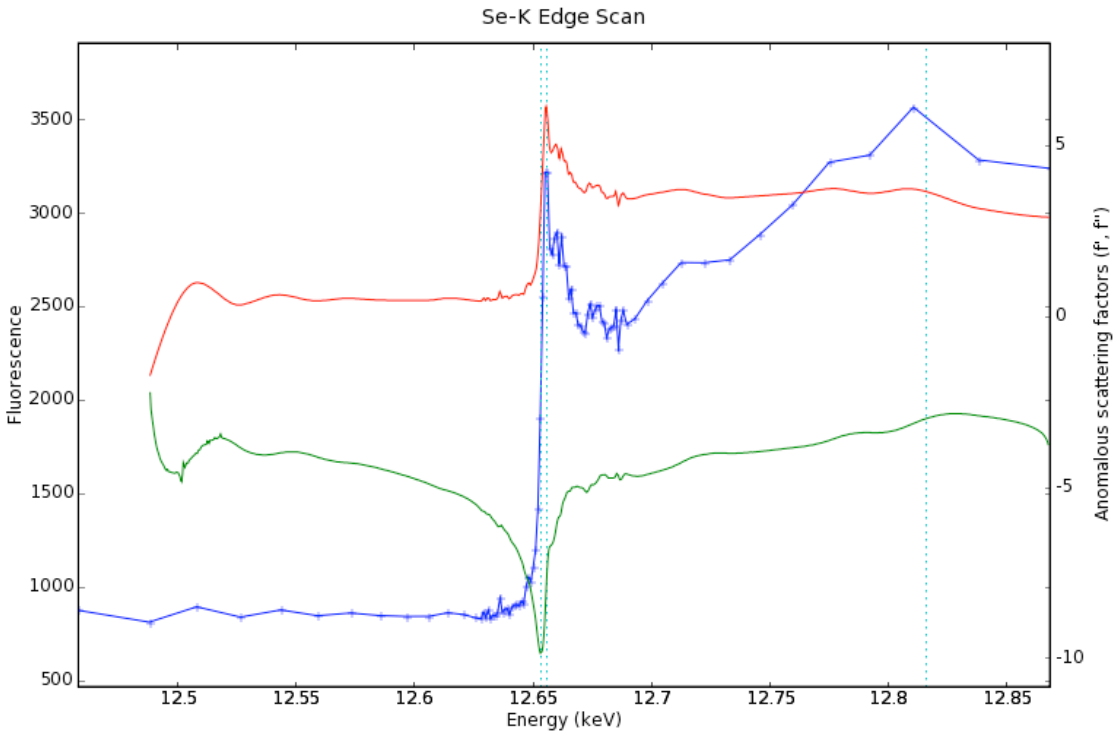


Figure 3.21: Fluorescence scan of the Se-K edge. A fluorescence scan at the Se-K edge of the Se-Met *D. melanogaster* MOF chromodomain (371-440) L423M crystals (blue) indicates Se-Met incorporation. The real (f') and imaginary (f'') components of the anomalous scattering factors are shown in red and green, respectively.

Based upon comparison of the processed data from the respective data sets, the Se_2 data set (peak and inflection) was concluded to be of the highest quality, due to a Chi^2 approximately equal to one, the lowest R_{symm}, highest resolution, and good completeness. It should be noted, however, that the redundancy was quite low (3.9). A high redundancy prevents erroneous rejection of a reflection (hkl) as an outlier.

All the data sets collected were analyzed to ascertain the strength of the anomalous signal, which solution of the phase problem and subsequent resolution of the crystal structure relies upon. When Chi^2 is plotted against resolution, Chi^2 values greater than one and the exhibition of resolution dependence is a strong indicator of an anomalous signal whereas Chi^2 approximately equal to one and no resolution dependence indicates that there is no useful anomalous signal. The data sets collected all exhibit Chi^2 close to one and showed little resolution dependency. Also, the Chi^2 plot does not change if the Friedel pairs are treated as equivalent. To further analyze the quality of the anomalous signal, the SOLVE software (Terwilliger and Berendzen, 1999) was used to analyze the correlation of the anomalous signal between the peak and inflection data sets. SOLVE was run for all the Se data sets but the majority of the focus was on the Se_2 data set as this was determined to be of the highest quality. As a general guideline, data should be cut-off where the correlation drops below 0.3 and a good data set should have an overall correlation of 0.7. The correlation on the anomalous signal between the peak and inflection indicated that the data should be cut-off around 3.3 Å (Table 3.7).

The crystallographic software SOLVE (Terwilliger and Berendzen, 1999) was used to calculate the Patterson map and the derived position of the heavy atoms. Initial electron density maps were then calculated by SOLVE using the calculated phases for the heavy atoms. SOLVE analyses the non-randomness of the native Fourier map. Maps with distinct solvent regions typically have a high standard deviation (SD) of the local rms electron density (Terwilliger, 2003). For the maps calculated using the Se_2 data, SOLVE reported a local rms of 0.25, which is not very high. Maps should also have a high correlation of local rms electron density at neighboring boxes. SOLVE reports a poor solution as having a local rms of 0.11. The calculated local rms for this solution was 9.55E-02. The overall Z score for the top solution found using SOLVE, which factors in the score from the native Fourier, the difference Patterson, the cross-validation Fourier, and the mean figure of merit was 9.22.

Table 3.6: HKL2000 Results. The individual Se-Met MAD data sets were indexed, integrated, and scaled using HKL2000. Statistics for the scaled unmerged data sets are shown. Results for the Se_4 peak data set were calculated using 140 frames or (240) frames.

	Se_2		**Se_4**		**Se7**	
	peak	infl	peak	infl	peak	infl
No. Frames	240	240	140 (240)	140	80	80
Chi^2	1.124	1.093	2.6 (4.696)	2.77	1.07	1.034
$R_{sym\ (all\ hkl)}$	0.068	0.072	0.078 (0.116)	0.08	0.079	0.098
Resolution (Å)	50-2.6	50-2.6	50-3 (50-3)	50-3	50-2.8	50-2.8
Error scale factor	2.3	2.3	2.5 (2.5)	2.5	1.4	1.4
Completeness (all hkl)	100	99.9	99.9	99.9	90.5	89
Redundancy (all hkl)	3.9	3.9	4.4 (7.3)	4.4	5.4	5.1

Table 3.7: Correlation coefficients in the anomalous signal between the peak and inflection data set for Se_2.

Resolution Shell	**Correlation**
6	0.83
4.5	0.71
4.2	0.58
3.98	0.55
3.75	0.5
3.6	0.43
3.45	0.38
3.3	0.28
3.15	0.16
3	0.23
All	0.46

This was the highest Z score calculated for a solution from the five data sets. The algorithm in SOLVE for finding, evaluating and selecting heavy-atom solutions is iterative. SOLVE stops searching for solutions once a solution with a Z score above 10 has been identified (Terwilliger, 2002). Using this solution, RESOLVE (Terwilliger, 2003) was run. RESOLVE is a density modification program that uses the calculated structure factors, the proteins amino acid sequence, and the expected characteristics of an electron density map of a macromolecule to modify the electron density maps calculated by SOLVE and build an initial model. Analysis of the maps and the model generated by RESOLVE indicated that a solution had not been found; the density was not continuous, nor was the model, and in many instances the small fragments that had been built into the density did not make any stereochemical or conformational sense.

3.5.7 Data Collection and Data Processing of *D. melanogaster* MOF chromodomain Crystals Containing Br^- or I^- Anions.

Several avenues to introduce either iodide or bromide anions were investigated. The methods investigated involved crystal soaking, co-crystallization, chemical modification, and vapour diffusion. The only method that worked was crystal soaking. Wild type crystals were rapidly transferred from the mother liquor into a solution containing 1.5 M sodium citrate, 100 mM CHES pH 9.5, 0.5 M KI (or NaBr) before mounting and freezing under a stream of nitrogen at 114 K using a Kryoflex low temperature apparatus.

Initial data was collected at the Saskatchewan Structural Sciences Centre using a DX8 Proteum Copper (Cu) rotating anode diffractometer with a Kappa goniostat and Proteum 4K charge-coupled device (CCD) detector. Single isomorphous replacement (SIR), Single wavelength anomalous dispersion (SAD), and single isomorphous replacement with anomalous scattering (SIRAS) were all investigated using the in-house iodinated and native data sets. SIR involves calculation of the position and phases of the heavy atom by constructing a isomorphous differences Patterson map. The structure factor of the heavy atom derivative (F_{PH}) is the vector sum of the heavy atom (F_H) and the native protein atoms in the crystal (F_P). Therefore a difference Patterson map can be calculated. SAD is a similar process, which constructs an anomalous differences Patterson map. SIRAS uses the combined isomorphous and anomalous data. Initial attempts at SIR gave very high R values, indicating that the two data sets were not indexed identically. In a hexagonal system, the definition of axes as a,b,c or b,a,-c are equally valid Figure 3.22 A). As a result, auto-indexing software may index using

either definition. Plotting the hk0 sections for the native and iodinated data indicated that this was the case in this instance (Figure 3.22 B), therefore the iodinated data was reindexed, changing the h,k,l to k,h,-l (Figure 3.22 C).

A total of 300° of data was collected over 600 frames, each with an oscillation of 0.5° around omega over a 1 min period. The diffraction images were indexed and processed using SaintPlus and ProScale (Bruker AXS Inc). The unit cell is isomorphous to the native data set. Data processing statistics are summarized in Table 3.8. SIR, SAD, and SIRAS did not give a clear solution as to the location of the incorporated heavy atoms.

Table 3.8: Summary of data processing for the iodinated *D. melanogaster* MOF chromodomain (371-440) crystal data set.

Res. Shell	nRef Obs	nRef Exp	Compl %	Redund	Rsym	Shell	Rfree	nRfree	I/ sigI	Shell	Chi^2	AnoI/ SigI	Shell	N (I+)	N (I-)
37.88 to 5.88	1060	1061	99.91	9.55	0.0665	0.0665	0.0678	51	9.97	9.97	0.99	6.51	6.51	5184	4935
4.63 to 4.05	2119	2120	99.95	9.78	0.0682	0.0695	0.0776	106	9.36	8.77	0.88	5.06	3.61	10644	10082
3.66 to 3.40	3172	3173	99.97	9.83	0.0703	0.0733	0.0757	155	9.22	8.95	0.91	4.61	3.71	16092	15095
3.19 to 3.03	4240	4241	99.98	9.87	0.0779	0.1012	0.0804	203	8.63	6.9	0.95	4.23	3.11	21633	20226
2.90	5293	5294	99.98	9.89	0.086	0.1326	0.0904	269	7.93	5.16	0.96	3.92	2.65	27090	25251
	6365	6366	99.98	9.89	0.0937	0.1602	0.0972	314	7.28	4.06	0.94	3.66	2.41	32652	30291
	7416	7417	99.99	9.77	0.1014	0.2113	0.1052	370	6.73	3.09	0.93	3.43	2.03	38118	34302
	8492	8493	99.99	9.61	0.1096	0.2881	0.1112	415	6.23	2.21	0.92	3.23	1.86	43702	37870

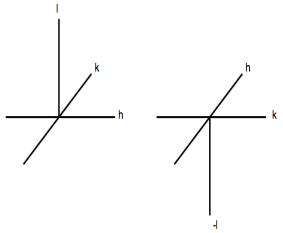
Maps calculated with SOLVE had a low correlation of local rms electron density at neighboring boxes and poor overall Z scores. The maps did not display any continuous electron density and RESOLVE failed to build a recognizable model. Fragments that were built into the density were small and did not make any stereochemical or conformational sense.

Further crystal soaking experiments were carried out in an attempt to collect better data at the Canadian Light Source (CLS) on the Canadian Macromolecular Facility 1 (CMCF1) 08ID-1 beamline. However, the procedure used for soaking the crystals is highly damaging to the crystals, based upon comparison of the diffraction. Data collected was not of high quality. A summary of the data collected is shown in Table 3.9.

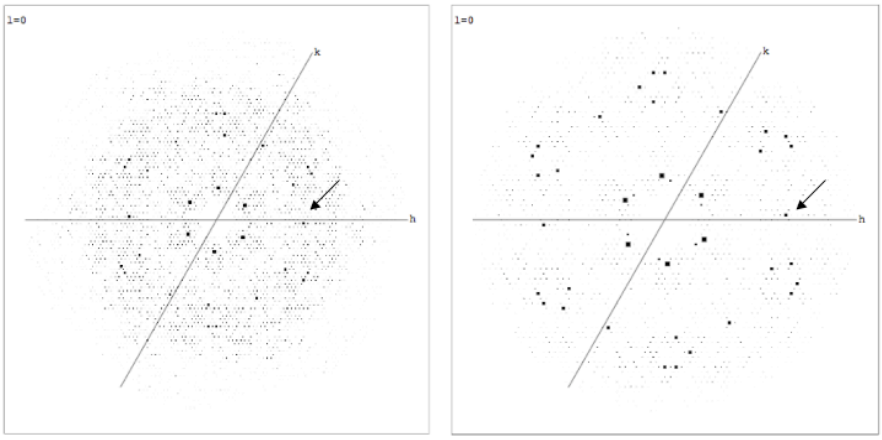
Table 3.9: Summary of crystal soaking data. Br1, Br3, Br4, Br5 and Br6 refer to the separate data sets collected on different crystals, soaked in the Br^- solution. Inflection data sets were not collected for Br1, Br4, Br5, and Br6 as the peak data set was of poor quality.

	Br1 peak	**Br3** peak	infl	**Br4** peak	**Br5** peak	**Br6** peak
No. Frames	90	150	150	120	120	120
Chi^2 (all hkl)	6.2	1.261	1.304	1.495	7.419	3.055
R_{sym} (all hkl)	0.107	0.11	0.101	0.1	0.165	0.145
Resolution (Å)	50-3.0	50-3.0	50-3.0	50-3.0	50-4.14	50-2.17
Completeness (all hkl)	97	99.9	99.9	99.8	99.6	75.7
Redundancy (all hkl)	5.8	4.9	4.9	7.4	7.3	5.8

A.

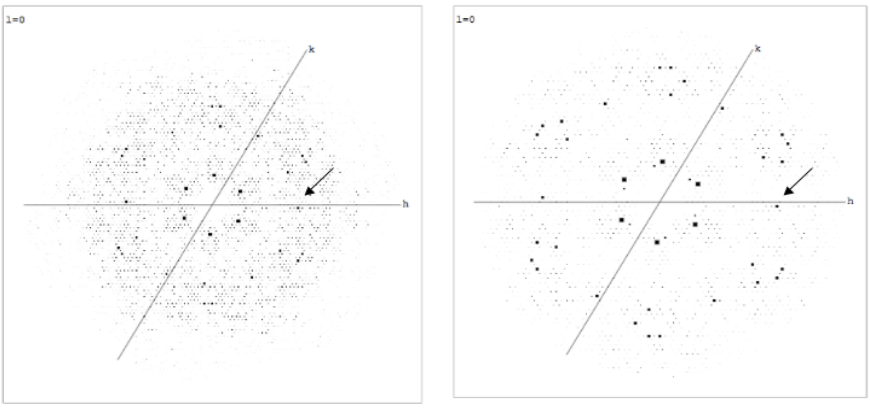


B.



Native Derivative

C.



Native Reindexed Derivative

Figure 3.22: Reindexing of the iodinated derivative data. **A.** The two axes hkl and k,h,-l are related by rotating k onto h. In a hexagonal system they are equivalent. **B.** hk0 section for the native and derivative data showing that they are not indexed identically. **C.** hk0 section for the native and derivative data after the derivative data has been reindexed.

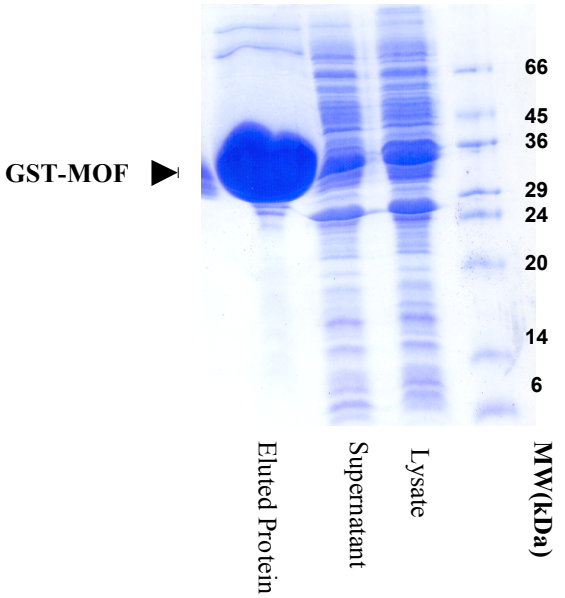
3.6 Cloning and Purification of the *H. sapiens* MOF Chromodomain (21-93)

The *H. sapiens* MOF (21-93) chromodomain has a theoretical pI and molecular weight of 5.31 and 8595.5 Da, respectively. When the eluted *H. sapiens* MOF (21-93) chromodomain was analyzed by SDS-PAGE (17% acrylamide:bis [37.5:1]) after GST affinity chromatography a large single band was observed at approximately 33 kDa, indicating that the construct expressed well. The GST affinity purification was effective in purifying the GST-MOF from the crude lysate as no other major bands were observed by SDS-PAGE (Figure 3.23 A). When digested with PreScission protease, two fragments are observed on the SDS-PAGE gel, as expected, with molecular weights consistent with the MOF (21-93) chromodomain (8 kDa) and GST (25 kDa) (Figure 3.23 B). After removal of the cleaved GST, the protein was further purified by anion exchange chromatography. Anion exchange was chosen due to the low pI of the protein (5.31). After anion exchange chromatography and SEC, the *H. sapiens* MOF chromodomain (21-93) was visibly homogenous, based upon SDS-PAGE (Figure 3.24 D). The MOF (21-93) chromodomain has a theoretical molecular weight of 8595.5 Da and had an apparent molecular weight of between 6 and 14 kDa, as determined by SDS-PAGE (17% acrylamide:bis [37.5:1]). The MOF (23-91) chromodomain eluted from the Superdex 75 column at a retention volume of 13.65 mL, which corresponds to a molecular weight of 14694 Da. The difference between the theoretical molecular weight (8595.5 Da) and the size of the protein eluting from the Superdex 75 column was significantly different. This may indicate that the MOF (23-91) chromodomain is a dimer in solution. The purified MOF (23-91) chromodomain was concentrated to approximately 18 mg/mL, in a total volume of 200 μL, for a total yield of approximately 3.6 mg of protein from 2 L of culture.

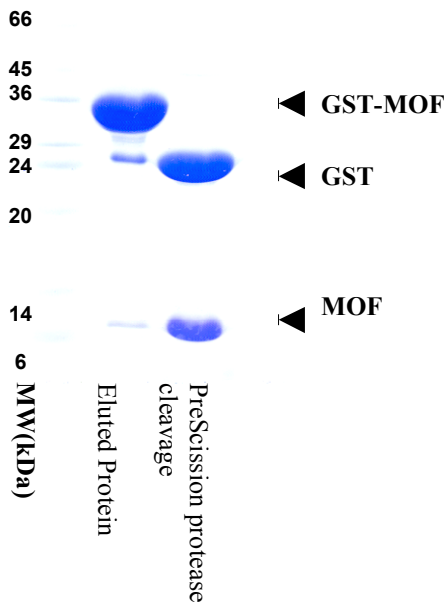
3.6.1 Vapour Diffusion Crystallization Screens

After concentrating, the MOF (23-91) was equilibrated into a buffer containing 20 mM Bis-Tris-propane, pH 7, 50 mM NaCl, 5 mM TCEP. Crystallization conditions were initially investigated using the Emerald Biosciences Wizard I and II and Hampton Research Index, PEG/Ion and Crystal Screen I and II commercial sparse matrix screens. Further crystal screens were set up screening pH versus sodium/potassium phosphate, lithium sulfate, or ammonium sulfate based upon precipitate observed in similar conditions from the commercial screens. No conditions were identified that yielded observable crystals.

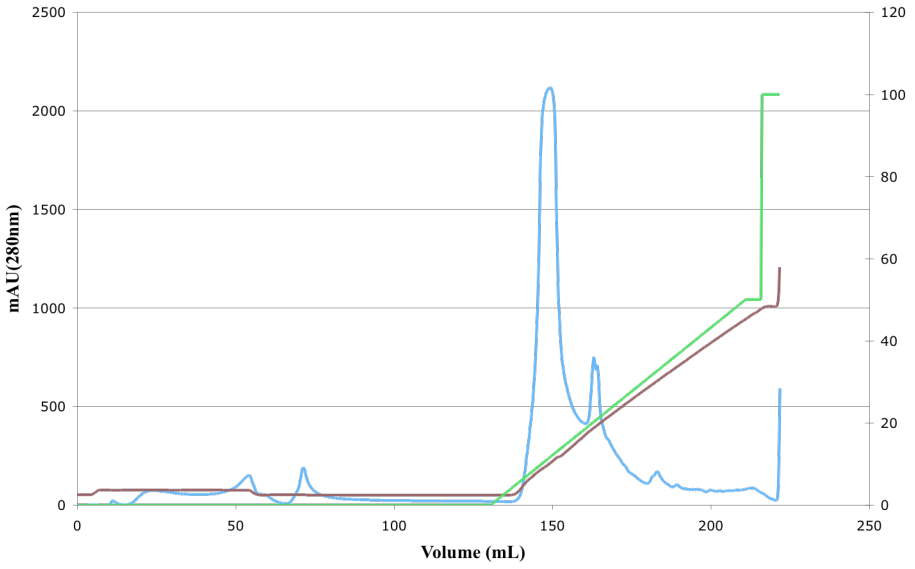
A.


GST-MOF
66
45
36
29
24
20
14
6
Eluted Protein
Supernatant
Lysate
MW(kDa)

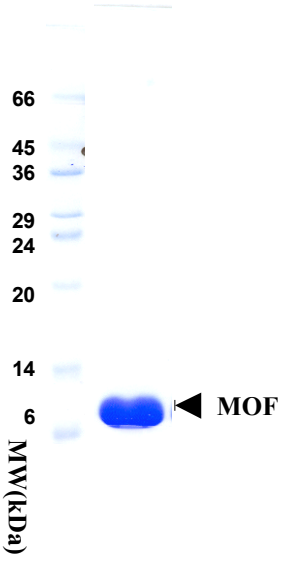
B.


66
45
36
29
24
20
14
6
GST-MOF
GST
MOF
MW(kDa)
Eluted Protein
PreScission protease cleavage

C.


2500
2000
1500
1000
500
0
mAU(280nm)
120
100
80
60
40
20
0
%B
0
50
100
150
200
250
Volume (mL)

D.


66
45
36
29
24
20
14
6
MOF
MW(kDa)

E.

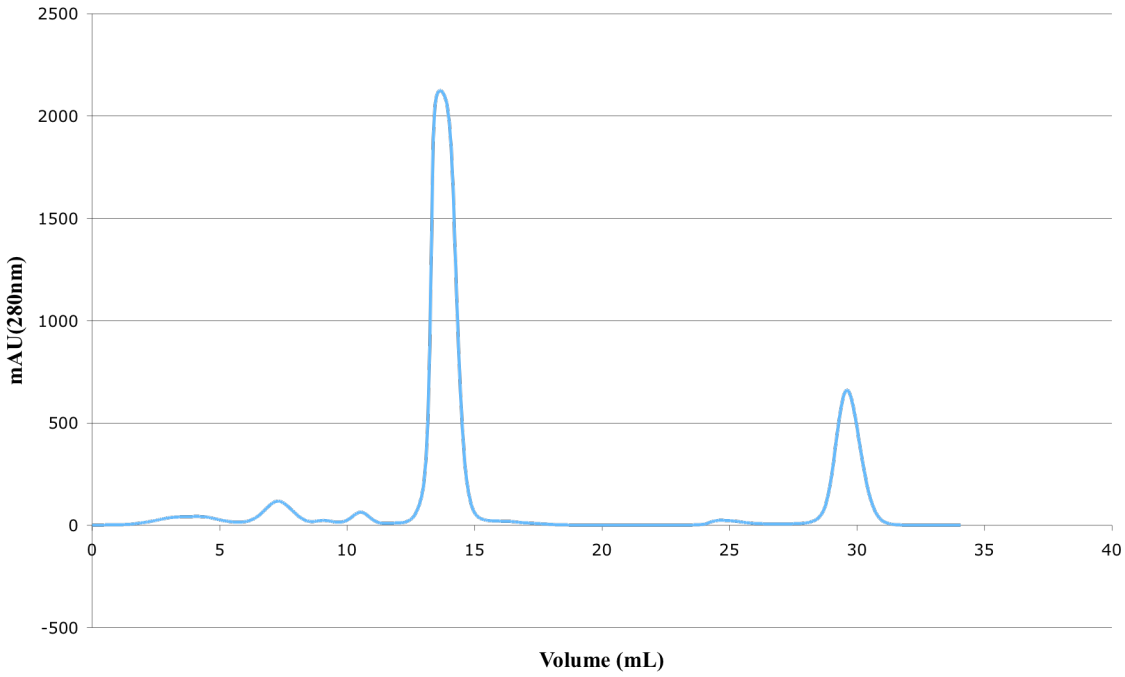


Figure 3.23: Purification of the *H. sapiens* MOF chromodomain (21-93). **A.** SDS-PAGE (17% acrylamide:bis [37.5:1]) of the eluted GST-MOF after GST affinity purification. Molecular weight markers are shown. **B.** SDS-PAGE (17% acrylamide:bis [37.5:1]) of the GST-MOF after cleavage with PreScission protease. Molecular weight markers are shown. **C.** Anion exchange chromatogram of the MOF chromodomain. The cleaved MOF chromodomain was loaded onto a SourceS cation exchange column at 1 mL/min. The protein was bound and washed with buffer A (see section 2.3.13.1) prior to a gradient elution with buffer B. Absorbance (280 nm) is shown in blue (mAU is absorbing units x 1000), conductivity (indicating salt concentration) in brown (mS/cm), and concentration (percentage of buffer B) is shown in green (%B). **D.** SDS-PAGE (17% acrylamide:bis [37.5:1]) of the eluted MOF chromodomain peak, collected in 1 mL. **E.** Size exclusion chromatography of the MOF chromodomain after pooling and concentration of the fractions from cation exchange. 500 µL was loaded onto a Superdex 75 column at a flow rate of 0.5 mL/min. The eluate was collected in 1 mL fractions.

4.0 Discussion

4.1 Chromodomains of the *D. melanogaster* Dosage Compensation Complex

At the outset of this research the precise functions of the MSL3 and MOF chromodomains were unknown. Since then, the MSL3 chromodomain has been reported to bind methylated H3K36Me$_3$ (Larschan *et al*., 2007; Sural *et al*., 2008). It is suggested that recognition of the H3K36Me$_3$ epigenetic mark is required for targeting and spreading of the DCC, in *cis* from the high affinity entry sites (Larschan *et al*., 2007; Sural *et al*., 2008). Corroborating this finding, the chromodomains of *S. cerevisiae* Eaf3 and *H. sapiens* MRG15, paralogues of the *D. melanogaster* MSL3 chromodomain, have also been demonstrated to bind H3K36Me$_3$ (Sun *et al*., 2008; Zhang *et al*., 2006). It was noted during initial attempts to purify the *D. melanogaster* MSL3 chromodomain (2-96) that the domain copurified with a nucleic acid. This conclusion was based upon the copurification of a species that was not visible on an SDS-PAGE gel stained with Coomassie Blue but was able to be ethanol precipitated, and stained with ethidium bromide when run on an agarose gel. Interaction of the MSL3 chromodomain with an oligonucleotide is in agreement with the findings by Akhtar, *et al*. (2000), who were the first to suggest that the MSL3 chromodomain binds roX RNA. The *D. melanogaster* dosage compensation complex contains two non-translated RNA molecules, roX1 and roX2. The notion that the MSL3 chromodomain (2-91) is sufficient for binding an oligonucleotide directly conflicts with previous findings that RNA binding is localized to residues 140-259 (Morales *et al*., 2005). Buscaino *et al*. (2006) suggest that the chromodomain and the polar region (140-259) act cooperatively to bind DNA. They demonstrated via electromobility shift assays that deletion of the chromodomain abrogated the ability of MSL3 to bind to DNA and mononucleosomes**.** Furthermore, they demonstrated that the chromodomain alone is able to bind to DNA and mononucleosome, but with reduced efficiently (Buscaino *et al*., 2006). It is likely that the *D. melanogaster* MSL3 chromodomain binds H3K36Me$_3$, and is involved in nucleotide binding. The MSL3 chromodomain is likely to play a key role in recruitment of MSL3 to the X chromosome (Buscaino *et al*., 2006; Larschan *et al*., 2007).

4.2 *D. melanogaster* MOF Chromodomain

4.2.1 Molecular Replacement: PHASER

The identification of conditions under which the *D. melanogaster* MOF chromodomain (371-440) reproducibly crystallized and the subsequent high quality of the diffraction data led to initial optimism that the solution of the crystal structure by molecular replacement using the *D. melanogaster* MOF chromodomain (367-454) NMR structure (Nielsen *et al*., 2005) would be relatively straight forward. However, the case proved problematic and the structure remains unsolved. This is likely due to the inherent problems of NMR models, the number of molecules in the asymmetric unit, the presence of non-crystallographic symmetry between monomers in the asymmetric unit, and the fact that the crystals have a hexagonal Bravais lattice. Each point is discussed below.

Molecular replacement relies upon strong structural similarity between the target structure and the search model. The use of an NMR structure in the solution of a crystal structure by molecular replacement was first demonstrated in 1987 (Brunger *et al*., 1987). While molecular replacement using NMR-derived atomic coordinates is possible, it is generally problematic because NMR structures inherently are based upon short distance restraints and have a low observation to parameter ratio. There is also difficulty in representing the relative reliability of atomic positions in an NMR model. Crystallographic models include *B* factor information, which applies weights to the relative atomic contribution of the scattering factors. NMR ensembles are used by PHASER to reflect the precision of atomic positions. The imprecision and lack of long distance restraints allows for the introduction of errors into an NMR structure, which may manifest as relatively large rms differences between an NMR and crystallographic model that share 100% sequence identity. The greater the structural difference between the search model and the target structure, the less likely molecular replacement will be successful. The quality of the search model can be improved by judicious editing of the atomic coordinates. In this case, the conformers of the ensemble were compared. Regions that showed variation were deleted and side chains that were disordered were truncated to the Cβ atom (Ala). It is also possible that the *D. melanogaster* MOF chromodomain NMR structure deviates significantly from the crystal structure, preventing solution of the rotation and translation function.

The asymmetric unit is the smallest part of a crystal to which crystallographic symmetry may be applied to generate the unit cell (Figure 4.1). The asymmetric unit may be a single molecule or multiple molecules, in the case of proteins.

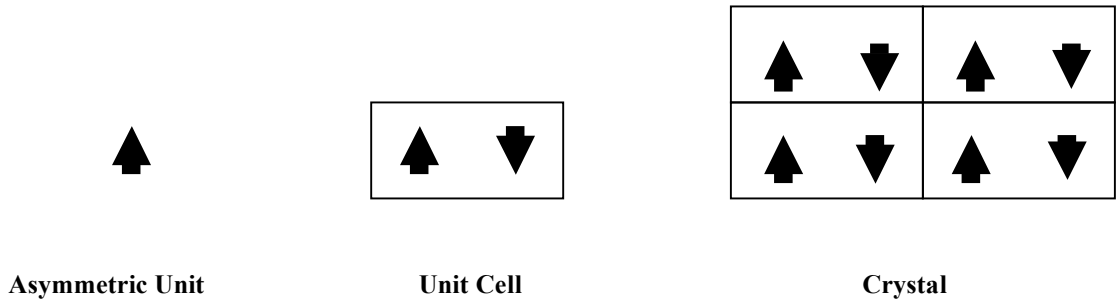


Figure 4.1: Graphical representation of the asymmetric unit, the unit cell, and the crystal structure and how they are related. In this instance the two asymmetric units in the unit cell are related by a two-fold rotation perpendicular to the plane of the page.

Molecular replacement calculates phases from the search model after it has been correctly positioned into the asymmetric unit. The rotation and translation functions applied to the search model are calculated by measuring the agreement between the calculated and observed Patterson functions, which are derived from the calculated and observed structure factor amplitudes. Solution of a crystal structure by molecular replacement becomes increasingly more difficult as the number of molecules in the asymmetric unit increases because the relative scattering contribution of a single molecule becomes smaller i.e. the signal to noise ratio decreases. In this instance, based upon the Matthew's coefficient (Matthews, 1968) there are probably 4 or 5 molecules in the asymmetric unit, therefore the relative scattering contribution of a single molecule is low and the signal that molecular replacement is searching for is small. This is further confounded by the crystals having a hexagonal Bravais lattice, as PHASER is searching six asymmetric units, compared to just one in the case of a monoclinic unit cell. A hexagonal Bravais lattice therefore further reduces the signal to noise ratio. High

non-crystallographic symmetry between molecules in the asymmetric unit, which is likely in this case further reduces the signal to noise ratio, making the translation search more difficult. PHASER uses a tree search with a pruning method. In the instance of searching for multiple molecules in the asymmetric unit, this involves using placement of first molecule as a background for searching for the second molecule and so on. This effectively increases the signal to noise ratio for the second search. Therefore the relative score of the solution should increase drastically as more molecules are placed into the asymmetric unit. Poor solutions are pruned out of the search. The success of this method relies upon the top rotation and translation functions of the initial search. For the case of molecular replacement of the *D. melanogaster* MOF chromodomain, the number of molecules searched for varied up to four molecules. The relative scores for the initial solution were very low and none improved drastically when the second, third, and fourth molecules were placed. In numerous occasions, solutions with the highest values were rejected due to excessive intermolecular contacts.

In summary, failure of molecular replacement to solve the crystal structure of the *D. melanogaster* MOF chromodomain (371-440) is likely to be a combination of using an NMR model as the search model, the high number of molecules in the asymmetric unit, high non-crystallographic symmetry, and a hexagonal Bravais lattice.

4.2.2 Se-Met Incorporation and MAD

The high pH and sodium citrate (a bidentate chelater) concentration meant that the soaking of a heavy metal into the crystal was not possible due to irreversible precipitation. In order to introduce selenium as the anomalous scatterer via Se-Met incorporation, a Met residue first had to be mutated into the sequence as none existed in the wild type *D. melanogaster* MOF chromodomain sequence (371-440). A total of nine mutants were made, of which only one crystallized, near wild type conditions. This may be due to Met incorporation preventing intermolecular contacts essential to crystallization. The *D. melanogaster* MOF chromodomain is very small and therefore Met incorporation is highly likely to alter the surface properties of the protein. The *D. melanogaster* MOF chromodomain (371-440) L423M mutant crystallized near wild type conditions and the crystal morphology was identical to the wild type protein crystals. It was noted that the crystals took one to two days longer to crystallize. The crystal initially had a undefined globular morphology with the distinct edges of the hexagonal prism

morphology resolving over approximately twenty-four hours. Neither the Met nor Se-Met version of the *D. melanogaster* MOF chromodomain (371-440) L423M mutant crystals diffracted as well as the wild type. The overall quality of the Se-Met derivative *D. melanogaster* MOF chromodomain (371-440) L423M mutant crystals diffraction was poor. While X-ray fluorescence scans confirmed the presence of Se-Met in the derivative crystal and an anomalous scattering signal was detected, it was not of high enough quality to solve the crystal structure by MAD. The poor anomalous signal could be due to a combination of the oxidation of the Se atoms, either during purification or diffraction, and/or disorder of the Se-Met residue in the crystal lattice, and/or the relative placement of the Se atom in the unit cell. Little can be done if the Se-Met residue is disordered. In order to reduce the likelihood of oxidation, purification of the Se-Met derivative crystal was carried out as fast as possible in the constant presence of a reducing agent. In order to reduce the likelihood of the crystals becoming damaged during diffraction the beam was attenuated and the diffraction experiment optimized to minimize the exposure of the crystal.

4.2.3 Iodide Incorporation and SIR

Parallel to the Se-Met MAD attempt, experiments were conducted to incorporate iodide (I^-) or bromide (Br^-) ions into the crystal lattice by crystal soaking. Soaking involved the rapid transfer of the crystal from its mother liquor into a buffer containing KI or NaBr. Initially experiments saw the crystals visibly crack and dissolve within seconds of transfer. This is likely due to the salting-in effect of the salt, increasing the solubility of the protein. To overcome this, the sodium citrate concentration and the pH in the soaking buffer had to be increased to sufficiently decrease the solubility of the protein, preventing cracking and dissolving. Based upon diffraction of the soaked crystals, the soaking was still likely to be damaging. A large peak at the Br^- absorption edge in the X-ray fluorescence spectrum, collected at the CLS confirmed incorporation of the halide, the data was poor and was likely not of sufficient quality to solve the crystal structure. The data was poor due to the likely damage of the crystal caused by soaking. Physical movement and soaking of the crystal along with incorporation of the halide ions could be interrupting crystal contacts causing large defects in the crystal lattice, which ultimately caused poor diffraction. Attempts to co-crystallize the protein in the presence of KI or NaBr also likely failed due to the salting-in effect on the protein.

4.3 Future Directions

The research presented outlines the cloning, expression and purification of the *D. melanogaster* MOF and MSL3 chromodomains and their *H. sapiens* orthologues, along with the *S. pombe* SWI6 chromodomain.

Crystallization conditions have been determined for the native *D. melanogaster* MOF (371-440) chromodomain. Future work will focus on crystallizing an isomorphous derivative containing a heavy atom, and collection of data of sufficient quality to allow for solution of the *D. melanogaster* MOF chromodomain crystal structure. Crystallization conditions for the other three constructs can also be investigated. Parallel to solving the crystal structure of the *D. melanogaster* MOF chromodomain, the molecular target needs also to be investigated to shed light on the role of the MOF chromodomain in the *D. melanogaster* dosage compensation complex. Binding studies between recombinant *D. melanogaster* MOF chromodomain and peptides representative of post-translationally modified histone tails could involve surface plasmon resonance, isothermal calorimetry, tryptophan fluorescence, and NMR techniques. These techniques can also be applied to structural and binding studies of the *H. sapiens* MOF and MSL3, *S. pombe* SWI6, and the *D. melanogaster* MSL3 chromodomain.

5.0 References

Aasland, R., Stewart, A.F., and Gibson, T. (1996). The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. Trends Biochem Sci *21*, 87-88.

Akhtar, A., and Becker, P.B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Mol Cell *5*, 367-375.

Akhtar, A., and Becker, P.B. (2001). The histone H4 acetyltransferase MOF uses a C2HC zinc finger for substrate recognition. EMBO Rep *2*, 113-118.

Akhtar, A., Zink, D., and Becker, P.B. (2000). Chromodomains are protein-RNA interaction modules. Nature *407*, 405-409.

Alekseyenko, A.A., Larschan, E., Lai, W.R., Park, P.J., and Kuroda, M.I. (2006). High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. Genes Dev *20*, 848-857.

Allard, S., Utley, R.T., Savard, J., Clarke, A., Grant, P., Brandl, C.J., Pillus, L., Workman, J.L., and Cote, J. (1999). NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. EMBO J *18*, 5108-5119.

Anand, K., Pallares, I., Valnickova, Z., Christensen, T., Vendrell, J., Wendt, K.U., Schreuder, H.A., Enghild, J.J., and Aviles, F.X. (2008). The crystal structure of thrombin-activable fibrinolysis inhibitor (TAFI) provides the structural basis for its intrinsic activity and the short half-life of TAFIa. J Biol Chem *283*, 29416-29423.

Armstrong, J.A., Papoulas, O., Daubresse, G., Sperling, A.S., Lis, J.T., Scott, M.P., and Tamkun, J.W. (2002). The Drosophila BRM complex facilitates global transcription by RNA polymerase II. EMBO J *21*, 5245-5254.

Arnold, A.P., Itoh, Y., and Melamed, E. (2008). A Birds-Eye View of Sex Chromosome Dosage Compensation. Annu Rev Genomics Hum Genet.

Badenhorst, P., Xiao, H., Cherbas, L., Kwon, S.Y., Voas, M., Rebay, I., Cherbas, P., and Wu, C. (2005). The Drosophila nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis. Genes Dev *19*, 2540-2545.

Bai, X., Larschan, E., Kwon, S.Y., Badenhorst, P., and Kuroda, M.I. (2007). Regional control of chromatin organization by noncoding roX RNAs and the NURF remodeling complex in Drosophila melanogaster. Genetics *176*, 1491-1499.

Bashaw, G.J., and Baker, B.S. (1995). The msl-2 dosage compensation gene of Drosophila encodes a putative DNA-binding protein whose expression is sex specifically regulated by Sex-lethal. Development *121*, 3245-3258.

Becker, P.B., and Horz, W. (2002). ATP-dependent nucleosome remodeling. Annu Rev Biochem *71*, 247-273.

Bell, O., Conrad, T., Kind, J., Wirbelauer, C., Akhtar, A., and Schubeler, D. (2008). Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in Drosophila melanogaster. Mol Cell Biol *28*, 3401-3409.

Belote, J.M., and Lucchesi, J.C. (1980). Male-specific lethal mutations of Drosophila melanogaster. Genetics *96*, 165-186.

Berndsen, C.E., Albaugh, B.N., Tan, S., and Denu, J.M. (2007). Catalytic mechanism of a MYST family histone acetyltransferase. Biochemistry *46*, 623-629.

Bernstein, E., and Hake, S.B. (2006). The nucleosome: a little variation goes a long way. Biochem Cell Biol *84*, 505-517.

Bhaumik, S.R., and Green, M.R. (2001). SAGA is an essential in vivo target of the yeast acidic activator Gal4p. Genes Dev *15*, 1935-1945.

Bopp, D., Schutt, C., Puro, J., Huang, H., and Nothiger, R. (1999). Recombination and disjunction in female germ cells of Drosophila depend on the germline activity of the gene sex-lethal. Development *126*, 5785-5794.

Botuyan, M.V., Lee, J., Ward, I.M., Kim, J.E., Thompson, J.R., Chen, J., and Mer, G. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell *127*, 1361-1373.

Bouazoune, K., Mitterweger, A., Langst, G., Imhof, A., Akhtar, A., Becker, P.B., and Brehm, A. (2002). The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. EMBO J *21*, 2430-2440.

Boudreault, A.A., Cronier, D., Selleck, W., Lacoste, N., Utley, R.T., Allard, S., Savard, J., Lane, W.S., Tan, S., and Cote, J. (2003). Yeast enhancer of polycomb defines global Esa1-dependent acetylation of chromatin. Genes Dev *17*, 1415-1428.

Brunger, A.T., Campbell, R.L., Clore, G.M., Gronenborn, A.M., Karplus, M., Petsko, G.A., and Teeter, M.M. (1987). Solution of a Protein Crystal Structure with a Model Obtained from NMR Interproton Distance Restraints. Science *235*, 1049-1053.

Buscaino, A., Kocher, T., Kind, J.H., Holz, H., Taipale, M., Wagner, K., Wilm, M., and Akhtar, A. (2003). MOF-regulated acetylation of MSL-3 in the Drosophila dosage compensation complex. Mol Cell *11*, 1265-1277.

Buscaino, A., Legube, G., and Akhtar, A. (2006). X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3. EMBO Rep *7*, 531-538.

Cai, Y., Jin, J., Tomomori-Sato, C., Sato, S., Sorokina, I., Parmely, T.J., Conaway, R.C., and Conaway, J.W. (2003). Identification of new subunits of the multiprotein mammalian TRRAP/TIP60-containing histone acetyltransferase complex. J Biol Chem *278*, 42733-42736.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P.*, et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Chadwick, B.P., and Willard, H.F. (2001). A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. J Cell Biol *152*, 375-384.

Chang, K.A., and Kuroda, M.I. (1998). Modulation of MSL1 abundance in female Drosophila contributes to the sex specificity of dosage compensation. Genetics *150*, 699-709.

Charier, G., Couprie, J., Alpha-Bazin, B., Meyer, V., Quemeneur, E., Guerois, R., Callebaut, I., Gilquin, B., and Zinn-Justin, S. (2004). The Tudor tandem of 53BP1: a new structural motif involved in DNA and RG-rich peptide binding. Structure *12*, 1551-1562.

Cheung, P., Tanner, K.G., Cheung, W.L., Sassone-Corsi, P., Denu, J.M., and Allis, C.D. (2000). Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. Mol Cell *5*, 905-915.

Cheung, W.L., Ajiro, K., Samejima, K., Kloc, M., Cheung, P., Mizzen, C.A., Beeser, A., Etkin, L.D., Chernoff, J., Earnshaw, W.C*., et al.* (2003). Apoptotic phosphorylation of histone H2B is mediated by mammalian sterile twenty kinase. Cell *113*, 507-517.

Chow, J.C., and Brown, C.J. (2003). Forming facultative heterochromatin: silencing of an X chromosome in mammalian females. Cell Mol Life Sci *60*, 2586-2603.

Clapier, C.R., Chakravarthy, S., Petosa, C., Fernandez-Tornero, C., Luger, K., and Muller, C.W. (2008). Structure of the Drosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer. Proteins *71*, 1-7.

Clapier, C.R., Langst, G., Corona, D.F., Becker, P.B., and Nightingale, K.P. (2001). Critical role for the histone H4 N terminus in nucleosome remodeling by ISWI. Mol Cell Biol *21*, 875-883.

Clapier, C.R., Nightingale, K.P., and Becker, P.B. (2002). A critical epitope for substrate recognition by the nucleosome remodeling ATPase ISWI. Nucleic Acids Res *30*, 649-655.

Cline, T.W. (1993). The Drosophila sex determination signal: how do flies count to two? Trends Genet *9*, 385-390.

Cohen, C., and Parry, D.A. (1990). Alpha-helical coiled coils and bundles: how to design an alpha-helical protein. Proteins *7*, 1-15.

Copps, K., Richman, R., Lyman, L.M., Chang, K.A., Rampersad-Ammons, J., and Kuroda, M.I. (1998). Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly. EMBO J *17*, 5409-5417.

Corona, D.F., Clapier, C.R., Becker, P.B., and Tamkun, J.W. (2002). Modulation of ISWI function by site-specific histone acetylation. EMBO Rep *3*, 242-247.

Costanzi, C., and Pehrson, J.R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599-601.

Cowell, I.G., and Austin, C.A. (1997). Self-association of chromo domain peptides. Biochim Biophys Acta *1337*, 198-206.

Craig, J.M. (2005). Heterochromatin--many flavours, common themes. Bioessays *27*, 17-28.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell *111*, 185-196.

Demakova, O.V., Kotlikova, I.V., Gordadze, P.R., Alekseyenko, A.A., Kuroda, M.I., and Zhimulev, I.F. (2003). The MSL complex levels are critical for its correct targeting to the chromosomes in Drosophila melanogaster. Chromosoma *112*, 103-115.

Deng, H., Bao, X., Cai, W., Blacketer, M.J., Belmont, A.S., Girton, J., Johansen, J., and Johansen, K.M. (2008). Ectopic histone H3S10 phosphorylation causes chromatin structure remodeling in Drosophila. Development *135*, 699-705.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491-496.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science *306*, 1571-1573.

Doyon, Y., Selleck, W., Lane, W.S., Tan, S., and Cote, J. (2004). Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. Mol Cell Biol *24*, 1884-1896.

Elfring, L.K., Deuring, R., McCallum, C.M., Peterson, C.L., and Tamkun, J.W. (1994). Identification and characterization of Drosophila relatives of the yeast transcriptional activator SNF2/SWI2. Mol Cell Biol *14*, 2225-2234.

Elgin, S.C. (1996). Heterochromatin and gene regulation in Drosophila. Curr Opin Genet Dev *6*, 193-202.

Fan, Y., Nikitina, T., Morin-Kensicki, E.M., Zhao, J., Magnuson, T.R., Woodcock, C.L., and Skoultchi, A.I. (2003). H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. Mol Cell Biol *23*, 4559-4572.

Fingerman, I.M., Li, H.C., and Briggs, S.D. (2007). A charge-based interaction between histone H4 and Dot1 is required for H3K79 methylation and telomere silencing: identification of a new trans-histone pathway. Genes Dev *21*, 2018-2029.

Flanagan, J.F., Mi, L.Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature *438*, 1181-1185.

Furuhashi, H., Nakajima, M., and Hirose, S. (2006). DNA supercoiling factor contributes to dosage compensation in Drosophila. Development *133*, 4475-4483.

Gao, Y.G., Su, S.Y., Robinson, H., Padmanabhan, S., Lim, L., McCrary, B.S., Edmondson, S.P., Shriver, J.W., and Wang, A.H. (1998). The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. Nat Struct Biol *5*, 782-786.

Garcia, S.N., Kirtane, B.M., Podlutsky, A.J., Pereira-Smith, O.M., and Tominaga, K. (2007). Mrg15 null and heterozygous mouse embryonic fibroblasts exhibit DNA-repair defects post exposure to gamma ionizing radiation. FEBS Lett *581*, 5275-5281.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M.*, et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

Gilfillan, G.D., Straub, T., de Wit, E., Greil, F., Lamm, R., van Steensel, B., and Becker, P.B. (2006). Chromosome-wide gene-specific targeting of the Drosophila dosage compensation complex. Genes Dev *20*, 858-870.

Govind, C.K., Yoon, S., Qiu, H., Govind, S., and Hinnebusch, A.G. (2005). Simultaneous recruitment of coactivators by Gcn4p stimulates multiple steps of transcription in vivo. Mol Cell Biol *25*, 5626-5638.

Greenberg, A.J., Yanowitz, J.L., and Schedl, P. (2004). The Drosophila GAGA factor is required for dosage compensation in males and for the formation of the male-specific-lethal complex chromatin entry site at 12DE. Genetics *166*, 279-289.

Grienenberger, A., Miotto, B., Sagnier, T., Cavalli, G., Schramke, V., Geli, V., Mariol, M.C., Berenger, H., Graba, Y., and Pradel, J. (2002). The MYST domain acetyltransferase Chameau functions in epigenetic mechanisms of transcriptional repression. Curr Biol *12*, 762-766.

Gu, W., Wei, X., Pannuti, A., and Lucchesi, J.C. (2000). Targeting the chromatin-remodeling MSL complex of Drosophila to its sites of action on the X chromosome requires both acetyl transferase and ATPase activities. EMBO J *19*, 5202-5211.

Gupta, A., Sharma, G.G., Young, C.S., Agarwal, M., Smith, E.R., Paull, T.T., Lucchesi, J.C., Khanna, K.K., Ludwig, T., and Pandita, T.K. (2005). Involvement of human MOF in ATM function. Mol Cell Biol *25*, 5292-5305.

Gupta, V., Parisi, M., Sturgill, D., Nuttall, R., Doctolero, M., Dudko, O.K., Malley, J.D., Eastman, P.S., and Oliver, B. (2006). Global analysis of X-chromosome dosage compensation. J Biol *5*, 3.

Hake, S.B., and Allis, C.D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". Proc Natl Acad Sci U S A *103*, 6428-6435.

Hamada, F.N., Park, P.J., Gordadze, P.R., and Kuroda, M.I. (2005). Global regulation of X chromosomal genes by the MSL complex in Drosophila melanogaster. Genes Dev *19*, 2289-2294.

Hamiche, A., Kang, J.G., Dennis, C., Xiao, H., and Wu, C. (2001). Histone tails modulate nucleosome mobility and regulate ATP-dependent nucleosome sliding by NURF. Proc Natl Acad Sci U S A *98*, 14316-14321.

Hirota, T., Lipp, J.J., Toh, B.H., and Peters, J.M. (2005). Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. Nature *438*, 1176-1180.

Hong, L., Schroth, G.P., Matthews, H.R., Yau, P., and Bradbury, E.M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. J Biol Chem *268*, 305-314.

Huang, Y., Fang, J., Bedford, M.T., Zhang, Y., and Xu, R.M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science *312*, 748-751.

Izzo, A., Regnard, C., Morales, V., Kremmer, E., and Becker, P.B. (2008). Structure-function analysis of the RNA helicase maleless. Nucleic Acids Res *36*, 950-962.

Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080-2083.

Jin, Y., Wang, Y., Walker, D.L., Dong, H., Conley, C., Johansen, J., and Johansen, K.M. (1999). JIL-1: a novel chromosomal tandem kinase implicated in transcriptional regulation in Drosophila. Mol Cell *4*, 129-135.

Joshi, A.A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Mol Cell *20*, 971-978.

Kelley, R.L., Meller, V.H., Gordadze, P.R., Roman, G., Davis, R.L., and Kuroda, M.I. (1999). Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell *98*, 513-522.

Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Mol Cell Biol *25*, 3305-3316.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C.*, et al.* (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol Cell Biol *23*, 4207-4218.

Kuo, M.H., Zhou, J., Jambeck, P., Churchill, M.E., and Allis, C.D. (1998). Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. Genes Dev *12*, 627-639.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lane, M.C., O'Toole, P.W., and Moore, S.A. (2006). Molecular basis of the interaction between the flagellar export proteins FliI and FliH from Helicobacter pylori. J Biol Chem *281*, 508-517.

Langst, G., and Becker, P.B. (2001). Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. J Cell Sci *114*, 2561-2568.

Larschan, E., Alekseyenko, A.A., Gortchakov, A.A., Peng, S., Li, B., Yang, P., Workman, J.L., Park, P.J., and Kuroda, M.I. (2007). MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. Mol Cell *28*, 121-133.

Larsson, J., and Meller, V.H. (2006). Dosage compensation, the origin and the afterlife of sex chromosomes. Chromosome Res *14*, 417-431.

Lee, C.G., Chang, K.A., Kuroda, M.I., and Hurwitz, J. (1997). The NTPase/helicase activities of Drosophila maleless, an essential factor in dosage compensation. Embo J *16*, 2671-2681.

Lee, J., Thompson, J.R., Botuyan, M.V., and Mer, G. (2008). Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. Nat Struct Mol Biol *15*, 109-111.

Legube, G., McWeeney, S.K., Lercher, M.J., and Akhtar, A. (2006). X-chromosome-wide profiling of MSL-1 distribution and dosage compensation in Drosophila. Genes Dev *20*, 871-883.

Lever, M.A., Th'ng, J.P., Sun, X., and Hendzel, M.J. (2000). Rapid exchange of histone H1.1 on chromatin in living human cells. Nature *408*, 873-876.

Li, A., Maffey, A.H., Abbott, W.D., Conde e Silva, N., Prunell, A., Siino, J., Churikov, D., Zalensky, A.O., and Ausio, J. (2005a). Characterization of nucleosomes consisting of the human testis/sperm-specific histone H2B variant (hTSH2B). Biochemistry *44*, 2529-2535.

Li, B., Gogol, M., Carey, M., Lee, D., Seidel, C., and Workman, J.L. (2007a). Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. Science *316*, 1050-1054.

Li, F., Parry, D.A., and Scott, M.J. (2005b). The amino-terminal region of Drosophila MSL1 contains basic, glycine-rich, and leucine zipper-like motifs that promote X chromosome binding, self-association, and MSL2 binding, respectively. Mol Cell Biol *25*, 8913-8924.

Li, F., Schiemann, A.H., and Scott, M.J. (2008). Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the Drosophila MSL1/MSL2 complex. Mol Cell Biol *28*, 1252-1264.

Li, H., Fischle, W., Wang, W., Duncan, E.M., Liang, L., Murakami-Ishibe, S., Allis, C.D., and Patel, D.J. (2007b). Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. Mol Cell *28*, 677-691.

Lohe, A.R., Hilliker, A.J., and Roberts, P.A. (1993). Mapping simple repeated DNA sequences in heterochromatin of Drosophila melanogaster. Genetics *134*, 1149-1174.

Lucchesi, J.C. (1973). Dosage compensation in Drosophila. Annu Rev Genet *7*, 225-237.

Lucchesi, J.C. (1998). Dosage compensation in flies and worms: the ups and downs of X-chromosome regulation. Curr Opin Genet Dev *8*, 179-184.

Lucchesi, J.C., Kelly, W.G., and Panning, B. (2005). Chromatin remodeling in dosage compensation. Annu Rev Genet *39*, 615-651.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L). Nature *190*, 372-373.

Marin, I. (2003). Evolution of chromatin-remodeling complexes: comparative genomics reveals the ancient origin of "novel" compensasome genes. J Mol Evol *56*, 527-539.

Marley, J., Lu, M., and Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. J Biomol NMR *20*, 71-75.

Mateescu, B., England, P., Halgand, F., Yaniv, M., and Muchardt, C. (2004). Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. EMBO Rep *5*, 490-496.

Matthews, B.W. (1968). Solvent content of protein crystals. J Mol Biol *33*, 491-497.

Maurer-Stroh, S., Dickens, N.J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., and Ponting, C.P. (2003). The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem Sci *28*, 69-74.

McCoy, A.J., Grosee-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. Journal of Applied Crystallography *40*, 658-674.

Meller, V.H. (2003). Initiation of dosage compensation in Drosophila embryos depends on expression of the roX RNAs. Mech Dev *120*, 759-767.

Meller, V.H., Gordadze, P.R., Park, Y., Chu, X., Stuckenholz, C., Kelley, R.L., and Kuroda, M.I. (2000). Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in Drosophila. Curr Biol *10*, 136-143.

Meller, V.H., and Rattner, B.P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. EMBO J *21*, 1084-1091.

Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J.*, et al.* (2006). Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. Mol Cell *21*, 811-823.

Meyer, B.J., and Casson, L.P. (1986). Caenorhabditis elegans compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. Cell *47*, 871-881.

Min, J., Zhang, Y., and Xu, R.M. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev *17*, 1823-1828.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat Genet *37*, 1090-1097.

Morales, V., Regnard, C., Izzo, A., Vetter, I., and Becker, P.B. (2005). The MRG domain mediates the functional integration of MSL3 into the dosage compensation complex. Mol Cell Biol *25*, 5947-5954.

Morales, V., Straub, T., Neumann, M.F., Mengus, G., Akhtar, A., and Becker, P.B. (2004). Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. EMBO J *23*, 2258-2268.

Nguyen, D.K., and Disteche, C.M. (2006). Dosage compensation of the active X chromosome in mammals. Nat Genet *38*, 47-53.

Nielsen, P.R., Nietlispach, D., Buscaino, A., Warner, R.J., Akhtar, A., Murzin, A.G., Murzina, N.V., and Laue, E.D. (2005). Structure of the chromo barrel domain from the MOF acetyltransferase. J Biol Chem *280*, 32326-32331.

Nielsen, P.R., Nietlispach, D., Mott, H.R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A.G., Murzina, N.V., and Laue, E.D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature *416*, 103-107.

Oh, H., Park, Y., and Kuroda, M.I. (2003). Local spreading of MSL complexes from roX genes on the Drosophila X chromosome. Genes Dev *17*, 1334-1339.

Otwinowski, Z., and Minor, M. (1997). Processing of X-ray Diffraction Data Colelcted in Oscillation Mode. Methods in Enzymology *276*, 307-326.

Pardo, P.S., Leung, J.K., Lucchesi, J.C., and Pereira-Smith, O.M. (2002). MRG15, a novel chromodomain protein, is present in two distinct multiprotein complexes involved in transcriptional activation. J Biol Chem *277*, 50860-50866.

Paro, R., Strutt, H., and Cavalli, G. (1998). Heritable chromatin states induced by the Polycomb and trithorax group genes. Novartis Found Symp *214*, 51-61; discussion 61-56, 104-113.

Penalva, L.O., and Sanchez, L. (2003). RNA binding protein sex-lethal (Sxl) and control of Drosophila sex determination and dosage compensation. Microbiol Mol Biol Rev *67*, 343-359, table of contents.

Peterson, C.L., and Laniel, M.A. (2004). Histones and histone modifications. Curr Biol *14*, R546-551.

Pray-Grant, M.G., Daniel, J.A., Schieltz, D., Yates, J.R., 3rd, and Grant, P.A. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature *433*, 434-438.

Qiu, H., Hu, C., Yoon, S., Natarajan, K., Swanson, M.J., and Hinnebusch, A.G. (2004). An array of coactivators is required for optimal recruitment of TATA binding protein and RNA polymerase II by promoter-bound Gcn4p. Mol Cell Biol *24*, 4104-4117.

Rangasamy, D., Berven, L., Ridgway, P., and Tremethick, D.J. (2003). Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. EMBO J *22*, 1599-1607.

Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., and Bonner, W. (2002). Histone H2A variants H2AX and H2AZ. Curr Opin Genet Dev *12*, 162-169.

Reid, J.L., Moqtaderi, Z., and Struhl, K. (2004). Eaf3 regulates the global pattern of histone acetylation in Saccharomyces cerevisiae. Mol Cell Biol *24*, 757-764.

Richter, L., Bone, J.R., and Kuroda, M.I. (1996). RNA-dependent association of the Drosophila maleless protein with the male X chromosome. Genes Cells *1*, 325-336.

Sarcinella, E., Zuzarte, P.C., Lau, P.N., Draker, R., and Cheung, P. (2007). Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin. Mol Cell Biol *27*, 6457-6468.

Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138-141.

Schmid, M., Arib, G., Laemmli, C., Nishikawa, J., Durussel, T., and Laemmli, U.K. (2006). Nup-PI: the nucleopore-promoter interaction of genes in yeast. Mol Cell *21*, 379-391.

Schwab, D.J., Bruinsma, R.F., Rudnick, J., and Widom, J. (2008). Nucleosome switches. Phys Rev Lett *100*, 228105.

Scott, M.J., Pan, L.L., Cleland, S.B., Knox, A.L., and Heinrich, J. (2000). MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in Drosophila. EMBO J *19*, 144-155.

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., and Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nat Struct Biol *8*, 27-31.

Shimojo, H., Sano, N., Moriwaki, Y., Okuda, M., Horikoshi, M., and Nishimura, Y. (2008). Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain. J Mol Biol *378*, 987-1001.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844-847.

Smith, E.R., Allis, C.D., and Lucchesi, J.C. (2001). Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in Drosophila males. J Biol Chem *276*, 31483-31486.

Smith, E.R., Cayrou, C., Huang, R., Lane, W.S., Cote, J., and Lucchesi, J.C. (2005). A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. Mol Cell Biol *25*, 9175-9188.

Smith, E.R., Eisen, A., Gu, W., Sattah, M., Pannuti, A., Zhou, J., Cook, R.G., Lucchesi, J.C., and Allis, C.D. (1998). ESA1 is a histone acetyltransferase that is essential for growth in yeast. Proc Natl Acad Sci U S A *95*, 3561-3565.

Smith, E.R., Pannuti, A., Gu, W., Steurnagel, A., Cook, R.G., Allis, C.D., and Lucchesi, J.C. (2000). The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. Mol Cell Biol *20*, 312-318.

Sprangers, R., Groves, M.R., Sinning, I., and Sattler, M. (2003). High-resolution X-ray and NMR structures of the SMN Tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. J Mol Biol *327*, 507-520.

Srinivasan, S., Armstrong, J.A., Deuring, R., Dahlsveen, I.K., McNeill, H., and Tamkun, J.W. (2005). The Drosophila trithorax group protein Kismet facilitates an early step in transcriptional elongation by RNA Polymerase II. Development *132*, 1623-1635.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Straub, T., and Becker, P.B. (2007). Dosage compensation: the beginning and end of generalization. Nat Rev Genet *8*, 47-57.

Straub, T., Gilfillan, G.D., Maier, V.K., and Becker, P.B. (2005). The Drosophila MSL complex activates the transcription of target genes. Genes Dev *19*, 2284-2288.

Stuckenholz, C., Kageyama, Y., and Kuroda, M.I. (1999). Guilt by association: non-coding RNAs, chromosome-specific proteins and dosage compensation in Drosophila. Trends Genet *15*, 454-458.

Sudarsanam, P., and Winston, F. (2000). The Swi/Snf family nucleosome-remodeling complexes and transcriptional control. Trends Genet *16*, 345-351.

Sullivan, K.F. (2001). A solid foundation: functional specialization of centromeric chromatin. Curr Opin Genet Dev *11*, 182-188.

Sun, B., Hong, J., Zhang, P., Dong, X., Shen, X., Lin, D., and Ding, J. (2008). Molecular basis of the interaction of Saccharomyces cerevisiae Eaf3 chromo domain with methylated H3K36. J Biol Chem.

Sural, T.H., Peng, S., Li, B., Workman, J.L., Park, P.J., and Kuroda, M.I. (2008). The MSL3 chromodomain directs a key targeting step for dosage compensation of the Drosophila melanogaster X chromosome. Nat Struct Mol Biol.

Takami, Y., Nishi, R., and Nakayama, T. (2000). Histone H1 variants play individual roles in transcription regulation in the DT40 chicken B cell line. Biochem Biophys Res Commun *268*, 501-508.

Tanabe, M., Kouzmenko, A.P., Ito, S., Sawatsubashi, S., Suzuki, E., Fujiyama, S., Yamagata, K., Zhao, Y., Kimura, S., Ueda, T.*, et al.* (2008). Activation of facultatively silenced Drosophila loci associates with increased acetylation of histone H2AvD. Genes Cells.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol *14*, 1025-1040.

Terwilliger, T.C. (2002). Automated structure solution, density modification and model building. Acta Crystallogr D Biol Crystallogr *58*, 1937-1940.

Terwilliger, T.C. (2003). SOLVE and RESOLVE: automated structure solution and density modification. Methods Enzymol *374*, 22-37.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr *55*, 849-861.

Trievel, R.C., Rojas, J.R., Sterner, D.E., Venkataramani, R.N., Wang, L., Zhou, J., Allis, C.D., Berger, S.L., and Marmorstein, R. (1999). Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. Proc Natl Acad Sci U S A *96*, 8931-8936.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

Tsukiyama, T., and Wu, C. (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell *83*, 1011-1020.

Turner, B.M., Birley, A.J., and Lavender, J. (1992). Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. Cell *69*, 375-384.

Varga-Weisz, P.D., Wilm, M., Bonte, E., Dumas, K., Mann, M., and Becker, P.B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. Nature *388*, 598-602.

Vignali, M., Hassan, A.H., Neely, K.E., and Workman, J.L. (2000). ATP-dependent chromatin-remodeling complexes. Mol Cell Biol *20*, 1899-1910.

Wang, Y., Zhang, W., Jin, Y., Johansen, J., and Johansen, K.M. (2001). The JIL-1 tandem kinase mediates histone H3 phosphorylation and is required for maintenance of chromatin structure in Drosophila. Cell *105*, 433-443.

White, C.L., Suto, R.K., and Luger, K. (2001). Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. EMBO J *20*, 5207-5218.

Wierzbicki, A.T., and Jerzmanowski, A. (2005). Suppression of histone H1 genes in Arabidopsis results in heritable developmental defects and stochastic changes in DNA methylation. Genetics *169*, 997-1008.

Wirbelauer, C., Bell, O., and Schubeler, D. (2005). Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. Genes Dev *19*, 1761-1766.

Witt, O., Albig, W., and Doenecke, D. (1996). Testis-specific expression of a novel human H3 histone gene. Exp Cell Res *229*, 301-306.

Woodcock, C.L., Frado, L.L., and Rattner, J.B. (1984). The higher-order structure of chromatin: evidence for a helical ribbon arrangement. J Cell Biol *99*, 42-52.

Wu, C., Bassett, A., and Travers, A. (2007). A variable topology for the 30-nm chromatin fibre. EMBO Rep *8*, 1129-1134.

Xiao, H., Sandaltzopoulos, R., Wang, H.M., Hamiche, A., Ranallo, R., Lee, K.M., Fu, D., and Wu, C. (2001). Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. Mol Cell *8*, 531-543.

Yan, Y., Harper, S., Speicher, D.W., and Marmorstein, R. (2002). The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate. Nat Struct Biol *9*, 862-869.

Yang, D., Lu, H., Hong, Y., Jinks, T.M., Estes, P.A., and Erickson, J.W. (2001). Interpretation of X chromosome dose at Sex-lethal requires non-E-box sites for the basic helix-loop-helix proteins SISB and daughterless. Mol Cell Biol *21*, 1581-1592.

Yang, X.J. (2004). The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. Nucleic Acids Res *32*, 959-976.

Yochum, G.S., and Ayer, D.E. (2002). Role for the mortality factors MORF4, MRGX, and MRG15 in transcriptional repression via associations with Pf1, mSin3A, and Transducin-Like Enhancer of Split. Mol Cell Biol *22*, 7868-7876.

Zhang, P., Du, J., Sun, B., Dong, X., Xu, G., Zhou, J., Huang, Q., Liu, Q., Hao, Q., and Ding, J. (2006). Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic Acids Res *34*, 6621-6628.

Zlatanova, J., and Thakar, A. (2008). H2A.Z: view from the top. Structure *16*, 166-179.

Zraly, C.B., Marenda, D.R., and Dingwall, A.K. (2004). SNR1 (INI1/SNF5) mediates important cell growth functions of the Drosophila Brahma (SWI/SNF) chromatin remodeling complex. Genetics *168*, 199-214.